KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK (ESAT)
AFDELING PSI
Kasteelpark Arenberg 10, 3001 Heverlee (Belgium)

A robust statistical surface registration framework using implicit function representations

Application in craniofacial reconstruction

Promotoren:
Prof. Dr. ir. P. Suetens
Prof. Dr. ir. D. Vandermeulen

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Peter Claes

04 juni 2007

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK (ESAT)
AFDELING PSI
Kasteelpark Arenberg 10, 3001 Heverlee (Belgium)

A robust statistical surface registration framework using implicit function representations

Application in craniofacial reconstruction

Jury:
Prof. Dr. ir. P. Suetens (promotor)
Prof. Dr. ir. D. Vandermeulen (promotor)
Prof. Dr. ir. L. Van Gool
Prof. Dr. G. Willems
Prof. Dr. ir. P. Dutré
Prof. Dr. ir. J. Suykens
Prof. Dr. J. G. Clement
(University of Melbourne)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Peter Claes

ISBN 978-90-5682-809-7
U.D.C. 681.3*J2
Wett. dep. D/2007/7515/43

04 juni 2007

ISBN 978-90-5682-809-7
U.D.C. 681.3*J2
Wett. dep. D/2007/7515/43

Acknowledgements

Four and a half years ago I had the opportunity of employment at the Medical Image Computing (MIC) research laboratory located in the University Hospital Gasthuisberg, Leuven, Belgium. The goal of the proposed research was to develop an automatic computerized craniofacial reconstruction method. Financially, the conducted research over the past years was supported by the Flemish Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT, project IWT/GBOU/020195), by the K.U.Leuven (project /OF/GOA/2004/05) and by the Fund for Scientific Research - Flanders (FWO-Vlaanderen, project FWO/G.0258.02). The GBOU project and my research in particular was an interesting and fruitful collaboration between the electrical engineering department of the engineering faculty and the orthodontics and forensic odontology department of the dental science faculty.

This doctoral work is the product of team efforts and there is no I in team. Therefore I would like to thank all the people who contributed to this work either directly or indirectly. First of all, it would have been impossible for me to accomplish this dissertation without the guidance of prof. Paul Suetens and prof. Dirk Vandermeulen, being my promoters. Prof. Paul Suetens is the head of the medical image processing research group (MIC) and is besides a great scientist also a warm and helpful person. He always made the time to answer and to resolve various types of questions and problems. Furthermore, he has a keen interest in the ideas and suggestions of other people while making a research group related discission. Special thanks to prof. Dirk Vandermeulen who was my scientific mentor and most direct coach. He has a tremendous knowledge about various biometric, computer vision and image processing related methodologies and he is mainly responsible for the outcome of this doctoral thesis. Thank you for showing me the light concerning implicit functions. He also carried out the difficult task of proofreading every paper and report I wrote. Furthermore, I had the pleasure of personally knowing him and his family during various occasions besides work and I'm grateful for his and their friendship. I would hereby like to thank his family as well for enduring the time Dirk has spent doing work for me and for others in his spare time.

I want to thank my assessors, prof. Luc Van Gool and prof. Guy Willems. Prof. Luc Van Gool is mainly active and very successful within the field of computer vision and leads an industrial image processing research group

(VISICS). Thanks to him, I came in contact with fellow engineers working on similar but industrial instead of medical related research topics. One of them was Rik Fransens for whom I show a great admiration concerning his combined practical and theoretical skills. I want to thank him for introducing and explaining the Expectation-Maximization algorithm to me. Prof. Guy Willems is active in the field of orthodontics and forensic odontology and was responsible for the forensic relevancy and practicality of my work. He had a never ending motivation to stimulate all the project participants. I also want to thank Sven De Greef, who was a PhD student at the department of orthodontics and forensic odontology and who was my partner in crime over the past years. The acquisition of an elaborated database for craniofacial reconstruction purposes and the different conference trips made us spent a lot of time together, resulting in a good friendship.

I would also like to thank all the members of my jury for their interest and participation. Special thanks to all the volunteers who participated in my research to obtain an elaborate database of facial information.

The development of a craniofacial reconstruction technique is useless without the proper demand for such a method. Therefore, I thank the Belgium federal police for showing interest in my work. Especially many thanks to Peter Clauwaert and Ben Claes from the Forensic Facial Imaging (FFI) police unit. They were my direct contact persons within the federal police and both of them are experts in manual 2D and 3D craniofacial reconstructions. Thanks for all the discussions, useful tips and your interest concerning forensic reconstruction and identification.

I'm grateful to all my colleagues at the MIC laboratory. I would like to fondly remember prof. Johan Van Cleynenbreugel†. He stood at the base of the GBOU project, but unfortunately left us to soon. I will mostly remember his enthusiasm when I showed him my first craniofacial reconstruction result. Thanks to prof. Frederik Maes for introducing and explaining analytical based gradient descent optimizations to me. Thanks to Jeroen Hermans and Dieter Seghers for the numerous serious and less serious break conversations. Thanks to Dirk Loeckx for solving a wide range of queries concerning software, theoretical and practical issues. I'm grateful to Wouter Mollemans for providing me and Sven De Greef with the useful visualization tool, Forensim. I thank Dominique Delaere and Bart De Dobbelaer, for ensuring that the systems and networks were always up and running. Dominique was always ready to offer a solution to any problem and we had some nice conversations as well. I thank Annitta and Patricia, for their diligence, punctuality and persistence in ensuring that all the paperwork was up-to-date. I also thank my card playing partners during lunch break. Of course thanks to all the remaining team players: Bruyninckx Pieter, Ruijters Daniel, Lambelin Yves, Scheys Lennart, De Buck Stijn, Loubele Miet, Elen An, Roose Liesbet, D'Agostino Emiliano, Kiss Gabriel, De Groeve Pieter, Suetens Kevin, Verstreken Kris, Schutyser Filip, Van Delm Tinne, Peeters Gert, Srivastava Siddharth, Wang Qian and Wouters Jeroen.

Live to work or work to live? Being amongst friends, I was frequently

reminded that the latter option is the most promising one. Thanks to all my friends from Leuven, Antwerpen or elsewhere for the enjoyable and relaxing moments during my spare time. Special Thanks to Ellen, Anne, Kathy and Lesly for providing me with food, fuel and entertainment during the writing of this book.

Last but not least, I would like to thank my family. Special thanks, vader en moeder for teaching me the important values of life. Thank you for giving me all the chances and opportunities to get me this far. Thanks, Tim, Kim, Sam, Flore, Ann, Ronny, Cindy, Bomma Pirard and Bompa Claes for always being there for me and for showing a great interest in my work. Thanks to the rest of the family. Unfortunately, some important family members left me to soon to witness the result of my research. I dedicate this work to them, Bompa Pirard†, Bomma Claes† and Clarisse†.

Thanks, Bedankt, Merci, Danke.

Peter
Leuven, June 4, 2007.

Abstract

A robust statistical surface registration framework using implicit function representations. Application in craniofacial reconstruction

Surface registration refers to the establishment of the geometrical relationship between two or more surfaces. It is a key enabling technology to solve a wide range of 3D computer-based modeling and reconstruction problems. We present a robust, statistical, surface registration framework that optimizes an inter-surface distance measure over the parameters that define the geometric transformation. We use a memory efficient, continuous, smooth and analytical implicit surface representation, which has particularly interesting properties for registration purposes. Robustness against noise and outliers of the surface models during registration is obtained using a Maximum a Posteriori (MAP) formulation including a statistical inlier- and outlier-process model.

The main application of the proposed framework is forensic craniofacial reconstruction which aims at estimating the face associated to an unknown skull specimen for victim identification. Traditional methods are based on a manual 3D reconstruction by physically modeling a face on a skull replica with clay or plasticine. We present here a computer-based (semi-)automated reconstruction method that is both consistent and objective. It builds on a statistical model of the interrelationship between facial shape and soft-tissue thicknesses as measured on a predefined set of landmark points on the skull. The model can be considered as an elastic mask with elastic hemi-spherical dowels on the inside of the mask at the landmark locations, which is subsequently fitted to the skull such that the virtual dowels touch the skull. The elastic deformation of the model is based on the statistical likelihood as learned from a database of nearly 400 subjects. The whole procedure involves a number of registration tasks, for which the proposed framework is used. First, complete facial surfaces are assembled from partial surface patches using a rigid transformation model. Second, surfaces of different subjects are matched to each other using a non-rigid Thin Plate Spline based deformation model. The observed inter-individual deformations are statistically modeled via Principal Component Analysis, assuming a multivariate normal distribution. Finally, this facial

model (including the soft-tissue thicknesses) is aligned with the (virtual) skull specimen using a transformation model based on the learned statistical deformations. Quantitative and qualitative facial reconstruction evaluations are used to validate the skull registration framework, while comparing the statistical model and the implicit skull representation choices with more traditional craniofacial reconstruction choices. The validation results indicate that the proposed statistical craniofacial reconstruction approach is a very promising substitute for current procedures.

Robuuste statistisch gebaseerde oppervlakregistraties m.b.v impliciete functie voorstellingen. Toegepast op gezichtsreconstructies

Oppervlakkenregistratie duidt op de bepaling van de geometrische relatie tussen twee of meerdere oppervlakken. Het is een algemene technologie en strategie om een brede waaier van 3D computergebaseerde modellering en reconstructieproblemen op te lossen. Wij presenteren een robuuste statistisch gebaseerde oppervlakregistratie methodologie die de afstand tussen oppervlakken minimaliseert over de parameters die de geometrische transformatie definiëren. Hiervoor gebruiken we een geheugen efficiënte, continue, zacht verlopende en analytische impliciete oppervlakvoorstelling, welke interessante eigenschappen incorporeert voor registratiedoeleinden. Robuustheid tegen ruis of kleine fouten en onverklaarbaarheden of grote fouten wordt bekomen m.b.v een maximum a-posteriori formulering van de registratie, dewelke statistische verklaarbare en onverklaarbare proces modellen bevat.

De hoofdapplicatie van het voorgestelde registratieraamwerk is forensische gezichtsreconstructie welke oogt op het genereren van een gezichtsenveloppe schatting geassocieerd met een onbekende schedel voor identificatiedoeleinden. Traditionele manuele reconstructiemethoden bestaan erin om fysisch het gezicht te boetseren op de droge schedel m.b.v klei of plasticine. Wij presenteren een (semi-)automatische computergebaseerde gezichtsreconstructie methode die zowel consistent is als objectief. De techniek is gebaseerd op een statistisch gezichtsmodel betreffende de interrelatie van gezichtsvorm en zachte weefsel diktes gemeten in een voorgedefinieerde verzameling van anatomische merkpunten op de schedel. Het model kan bekeken worden als een elastisch masker met elastische halve bollen aan de binnenkant, dat vervolgens vervormd wordt naar de schedel toe zodat de halve bollen het schedeloppervlak raken. De elasticiteit van het masker en de bollen is gebaseerd op een statistische waarschijnlijkheid zoals die geleerd werd uit een databank met gezichtsschedel relaterende gegevens van bijna 400 levende personen. De hele procedure vergt een aantal registratietaken waarvoor we het voorgestelde registratieraamwerk gebruiken. Ten eerste worden volledige gezichtsoppervlakken geassembleerd van meerdere partiële gezichtsoppervlak opnames gebruikmakend van een rigide transformatiemodel. Ten tweede worden volledige gezicht-

soppervlakken van verschillende individuen met elkaar geregistreerd m.b.v een niet-rigide thin-plate-spline (TPS) transformatiemodel. De als dusdanig bekomen geobserveerde interindividuele vervormingen worden statistisch gemodelleerd m.b.v principale component analyse (PCA), waarbij we de veronderstelling aannemen dat de data normaal verdeeld is. Tot slot, wordt dit forensische gezichtsmodel (met zacht weefsel diktes) geregistreerd met een virtuele kopie van de onbekende schedel gebruikmakend van een transformatiemodel geleerd uit de statistische vervormingen. Kwantitatieve en kwalitatieve gezichtsreconstructie evaluaties worden gebruikt om de schedel registratietechniek te valideren. Hierbij vergelijken we het statistische gezichtsmodel met meer traditionele gezichtsmodellen. Bijkomend wordt de impliciete schedelvoorstelling vergeleken met een traditionele craniometrische puntgebaseerde schedelvoorstelling. De resultaten geven aan dat de voorgestelde statistisch gebaseerde reconstructietechniek een waardige vervanger is voor huidige procedures.

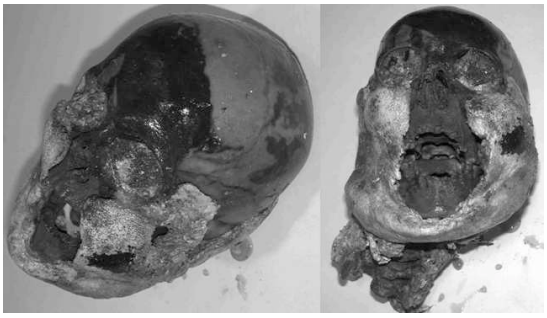
Nederlandse samenvatting

Robuuste statistisch gebaseerde oppervlakregistraties m.b.v impliciete functie voorstellingen. Toegepast op gezichtsreconstructies

0.1 Probleemstelling

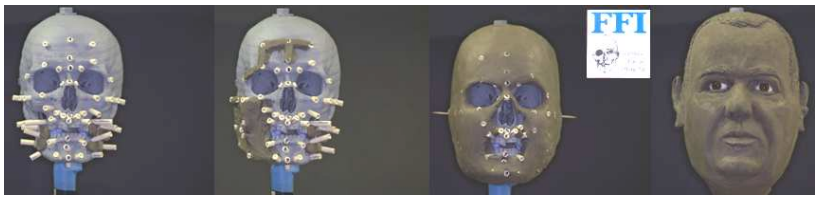
Waarom gezichtsreconstructie? Zondag ochtend, 7 uur, de hemel is helder en de zon komt op. Mr. Anoniem jogt langs het kanaal wanneer hij plotseling de schedelresten van een persoon tussen de bloemen in het struikgewas opmerkt. Hij belt onmiddellijk de autoriteiten met zijn mobiele telefoon en 20 minuten later arriveert de politie ter plaatse. Vooraleer eender welk onderzoek kan gestart worden is het belangrijk te achterhalen wie de persoon in het struikgewas is. Het antwoord op deze vraag wordt typisch verkregen m.b.v forensische identificatie technieken gebaseerd op het vergelijken van ante- en post-mortem data, tandheelkundige dossiers, X-rays of DNA. De hele procedure wordt minder evident wanneer men te maken heeft met humane resten in vergaande vorm van ontbinding zoals in figuur 0.1 waarbij elke relatie met een mogelijke identiteit ontbreekt. In deze omstandigheden kan een gezichtsreconstructie helpen om het onderzoek uit een impasse te halen. Een gezichtsreconstructie bestaat erin om een schatting van het gezicht van een individu te creëren op basis van de schedel op het tijdstip van overlijden. Verschillende 2D en 3D manuele en computergebaseerde gezichtsreconstructie methoden werden ontwikkeld over de jaren en allen zijn gebaseerd op de relatie tussen de zachte weefselenveloppe en het onderliggende harde weefsel of schedel structuur.

Van manuele naar computergebaseerde technieken Verscheidene 3D manuele methoden voor gezichtsreconstructie werden ontwikkeld en worden momenteel uitvoerig gebruikt in de praktijk. De technieken bestaan erin om fysisch het gezicht te boetseren op een kopie van de schedel m.b.v klei of plasticine (zie figuur 0.2). Deze methoden vragen echter een uitgebreide anatomische kennis en artistieke talenten zodat de uiteindelijke reconstructie



Figuur 0.1: Voorbeeld schedel in vergaande ontbinding.

resultaten eerder subjectief zijn (afhankelijk van de persoon die de reconstructie maakt). Darenboven vergen manuele reconstructies veel tijd waardoor ze vaak gelimiteerd zijn tot een enkele reconstructie. Computergebaseerde technieken zijn eerder consistent en objectief. Verder zijn de nodige berekeningen relatief snel uitgevoerd waardoor er meerdere reconstructies met verschillende modelleringassumpties (ouder, dikker,) verkregen kunnen worden. Met andere woorden, de ontwikkeling van computer software voor gezichtsreconstructies zou zeer interessant zijn voor verschillende gerechtsinstanties door sneller, eenvoudiger, efficiënter en objectiever meerdere reconstructies te genereren van een onbekend individu.



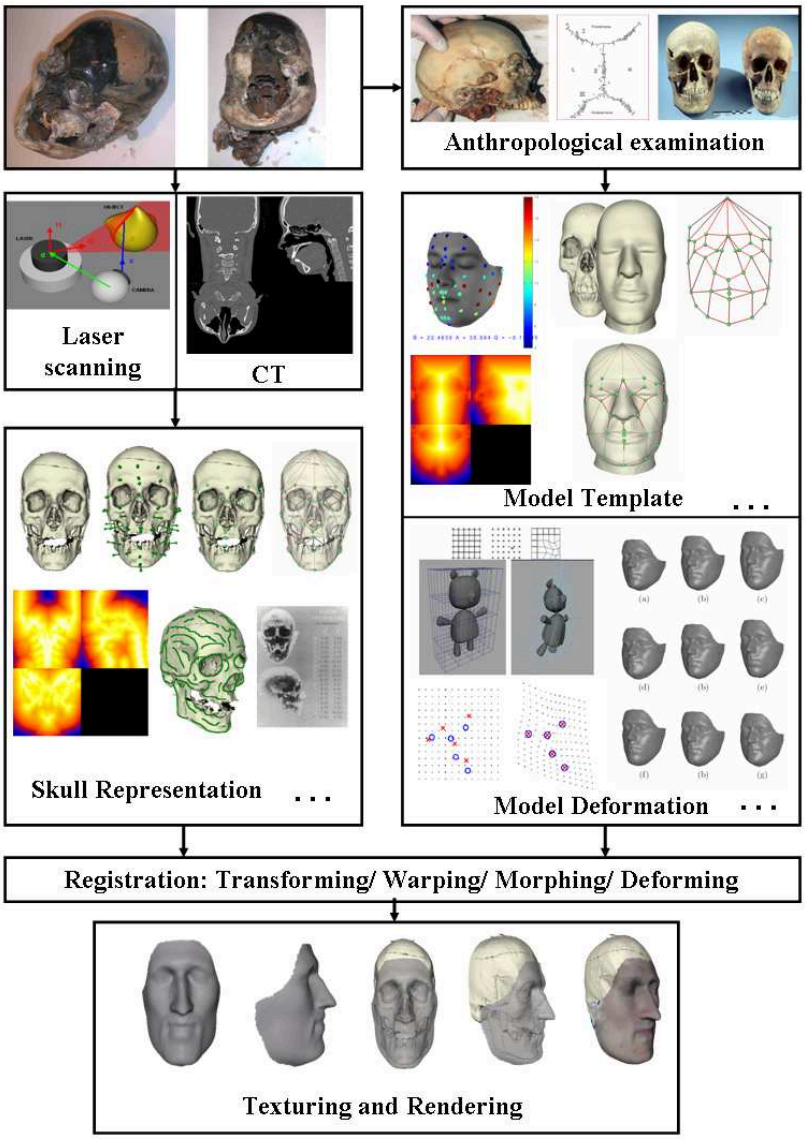
Figuur 0.2: Voorbeeld manuele reconstructiemethode uitgevoerd door de FFI afdeling van de Belgische federale Politie (Ben Claes)

Reconstructie raamwerk Computergebaseerde gezichtsreconstructie methoden gebruiken allen hetzelfde stramien of dezelfde strategie geïllustreerd in figuur 0.3. Een gezichtsmodel, gekozen op basis van schedeleigenschappen (afkomst, leeftijd, geslacht, ...), wordt vervormd of getransformeerd naar een virtuele kopie van de onbekende schedel, gebruikmakend van een bepaalde oppervlakvoorstelling (bvb. dichte punten wolk, beperkte anatomische schedel merkpunten, impliciete functie voorstelling, ...). Gelijkaardig aan een artiest die een manuele reconstructie uitvoert bevat het gezichtsmodel de vereiste kennis om het reconstructie probleem op te lossen en is typisch verkregen o.b.v relevante data van levende individuen uit een databank. Het gezichtsmodel bestaat uit een referentiegezicht of masker (bvb. een specifiek gezicht

uit de databank of een meer generisch gemiddelde gezicht) dat gezicht-schedel relaterende informatie bevat (bvb. zachte weefseldiktes, volledige schedeloppervlakken met bijbehorende gezichtsoppervlakken, ...) en wordt uitgebreid met een transformatiemodel om het masker te vervormen naar de schedel. De registratie van het model naar de schedel heeft als doel de bepaling van de geometrische relatie tussen het gezichtsmasker van het model en de schedel. De drijfveer van de registratie is een similariteitsmaat die de goedheid van overeenkomst tussen de schedel en het vervormde masker uitdrukt en die bepaald wordt op basis van de gekozen schedelvoorstelling en de gezicht-schedel relaterende informatie die vervat zit in het model. Het transformatiemodel en de similariteitsmaat worden gecombineerd in een doelfunctie waarvan het optimum overeenkomt met de ideale vervorming van het masker naar de schedel toe. Tijdens de registratie worden er transformatieparameters gezocht die het masker vervormen naar de schedel en die daardoor de maat van overeenkomst verbeteren tot er convergentie optreedt. Wanneer de ideale parameters gekend zijn kan het masker vervormd worden om een gezichtsreconstructie of een schatting van het onbekende gezicht te verkrijgen.

Wat ontbreekt er? Huidige computergebaseerde reconstructietechnieken zijn echter gelimiteerd in het gezichtsmodel dat ze gebruiken. Initieel wordt ofwel een enkelvoudig generisch ofwel een enkelvoudig specifiek meest gelijkaardig gezicht (in termen van schedel eigenschappen), gekozen als modelmasker. Vervolgens wordt een schatting van het onbekende gezicht verkregen door het modelmasker te vervormen, gebruikmakend van een generisch, zacht verlopend transformatiemodel (bvb. TPS gebaseerd). Meerdere reconstructies gebaseerd op verschillende schedeleigenschappen worden verkregen door andere gezichten uit de databank te kiezen als modelmasker. Twee grote tekortkomingen zijn echter aanwezig bij het gebruik van zulk enkelvoudig modelmasker in combinatie met een generisch transformatiemodel. Ten eerste wordt de reconstructie verkeerd beïnvloed door de keuze van referentiegezicht of modelmasker. Wanneer men gebruik maakt van een specifiek meest gelijkaardig gezicht blijven er specifieke gezichtskenmerken aanwezig in de reconstructie die afkomstig zijn van het referentiegezicht maar niet noodzakelijk horen bij het onbekende gezicht. Het gebruik van een gemiddeld gezicht als referentiegezicht of modelmasker resulteert in een gemiddelde reconstructie zonder enige vorm van specifieke gezichtskenmerken. Dit is niet noodzakelijk beter voor herkenningsdoeleinden. Ten tweede zijn generische transformatiemodellen niet gezichtsspecifiek, ze zijn enkel zacht verlopend. Er treden geen problemen op wanneer de verschillen tussen het referentiegezicht en de schedel klein zijn. Wanneer deze verschillen echter groot zijn is de nodige vervorming groot en het gevaar bestaat erin om gezichten te genereren met een lage gezichtswaardige waarschijnlijkheid. M.a.w het is niet eenvoudig om geldige en plausibele gezichtsreconstructies te bekomen gebruikmakend van een algemeen, niet gezichtspecifiek, transformatiemodel gecombineerd met een enkelvoudig referentiegezicht.

Het grootste verschil tussen bestaande computergebaseerde reconstructi-



Figuur 0.3: Algemene strategie voor computergebaseerde gezichtsreconstructie technieken

etechnieken is de databank waarvan vertrokken wordt om een gezichtsmodel met zekere gezicht-schedel relaterende informatie te extraheren. Dit beïnvloedt ook de keuze van schedeloppervlak voorstelling om het model naar schedelregistratie probleem op te lossen. Een ideale databank voor gezichtsreconstructie doeleinden bevat een groot aantal gegevens bestaande uit gezichtsoppervlakken en informatie die schedeloppervlakken met de gezichtsoppervlakken relateert. De gezichtsoppervlakken zijn bij voorkeur verkregen van personen in rechtstaande positie en de gezicht-schedel relaterende informatie bepaalt de keuze van de onbekende schedelvoorstelling. Een belangrijk punt betreffende de registratie van het gezichtsmodel naar de schedel is de aanwezigheid van kleine fouten of ruis en grote fouten of onverklaarbaarheden in de schedelvoorstelling. Buiten het feit dat een virtuele kopie van een schedel nooit exact is worden er extra kleine en grote fouten geïncorporeerd tijdens de opbouw van de gekozen schedeloppervlak voorstelling. Tot op heden, volgens het beste van onze kennis, wordt er geen robuuste (in termen van kleine en grote fouten) registratie methodologie gebruikt in huidige computergebaseerde gezichtsreconstructie technieken.

Een laatste punt betreffende computergebaseerde methoden is de nood aan validatie. Vanuit ingenieursstandpunt kan een ontwikkelde methode goed werken maar is waardeloos zonder een uitgebreide validatie. Een dergelijke validatie kan enkel een positieve bijdrage betekenen voor gezichtsreconstructie methoden in de praktijk. Een databank met gekende schedel - en gezichtsoppervlakken is noodzakelijk voor een goede evaluatie van een methode. Elke schedel in de databank kan gebruikt worden om een reconstructie te maken welke vervolgens vergeleken kan worden met het gekende gezicht van de schedel gebruikmakend van technieken voor gezichtsvergelijking.

0.2 Materialen en methoden

Robuuste statistisch gebaseerde gezichtsreconstructie In deze thesis presenteren wij een (semi-) automatische gezichtsreconstructie methode die consistent en objectief is. De techniek is gebouwd op een statistisch gezichtsmodel van de interrelatie tussen gezichtsvorm en zachte weefsel diktes gemeten in een voor gedefinieerde verzameling van merkpunten op de schedel. Het model kan bekeken worden als een elastisch masker met elastische halve bollen aan de binnenkant, dat vervolgens vervormd wordt naar de schedel toe zodat de halve bollen het schedeloppervlak raken. De elasticiteit van het modelmasker en de bollen is gebaseerd op een statistische waarschijnlijkheid zoals die geleerd werd uit een databank met gezicht-schedel relaterende gegevens van bijna 400 levende personen. Hiervoor zijn er een aantal deeltaken vereist.

1. Het verwerven van 3D gezichtsoppervlakken en diktemetingen van individuen in rechtstaande positie, gemeten over een voldoende diverse populatie, die vervolgens gestockeerd moeten worden in een databank samen met de leeftijd, BMI, en geslacht kenmerken van de individuen.

2. Het bepalen van puntovereenkomsten tussen de verschillende gezichtsoppervlakken van de individuen in de databank voor de constructie van een statistisch gezichtsmodel.

3. De statistische modellering van de populatieafhankelijke gecombineerde variatie en co-variatie van gezichtsvorm en zachte weefsel diktes, aangepast aan gegeven schedeleigenschappen (afkomst, geslacht, leeftijd, ...).

4. Het vervormen of registreren van dit gecombineerd statistisch gezichtsmodel naar de onbekende schedel.

5. Een uitgebreide validatie van de complete gezichtsreconstructie methode.

Robuust registratie raamwerk m.b.v impliciete oppervlakken De hele procedure (1-5) bevat een aantal registratie taken, wat refereert naar de bepaling van de geometrische relatie tussen twee of meerdere oppervlakken. Het is een algemene technologie en strategie om een brede waaier van 3D computergebaseerde modellering en reconstructie problemen op te lossen. Wij stellen een robuuste statistisch gebaseerde oppervlak registratie methodologie voor die de afstand tussen oppervlakken minimaliseert over de parameters die de geometrische transformatie definiëren. Hiervoor gebruiken we een geheugen efficiënte, continue, zacht verlopende en analytische impliciete oppervlakvoorstelling, welke interessante eigenschappen incorporeert voor registratie doeleinden. Robuustheid tegen ruis of kleine fouten en onverklaarbaarheden of grote fouten wordt bekomen m.b.v een maximum a-posteriori (MAP) formulering van de registratie, dewelke probabilistisch verklaarbare en onverklaarbare procesmodellen bevat. Dit registratieraamwerk wordt gehanteerd om de verschillende noodzakelijke registratie taken in de hele gezichtsreconstructie procedure op te lossen: in deeltaak (1) worden volledige gezichtsoppervlakken geassembleerd van meerdere partiële gezichtsoppervlak opnames gebruikmakend van een rigide transformatiemodel. In deeltaak (2) worden volledige gezichtsoppervlakken van verschillende individuen met elkaar geregistreerd m.b.v een niet-rigide thin-plate-spline (TPS) transformatiemodel. In deeltaak (3) worden de als dusdanig bekomen geobserveerde interindividuele vervormingen statistisch gemodelleerd m.b.v principale component analyse (PCA) onder de veronderstelling dat de data normaal verdeeld is. In een eerste instantie wordt dit statistische model geregistreerd met de gezichtsoppervlakken van de verschillende individuen in de databank d.m.v het geleerde gezichtsspecifieke statistische transformatiemodel om na te gaan of de TPS gebaseerde overeenkomstige punten kwalitatief voldoende en consistent zijn. Tot slot in deeltaak (4), wordt dit gezichtsmodel (plus de zachte weefsel diktes) geregistreerd met de virtuele kopie van de schedel door opnieuw gebruik te maken van de geleerde gezichtsspecifieke statistische vervormingen.

Het doel van deze thesis is om elke deeltaak binnen de procedure op te lossen, gebruikmakend van het voorgestelde registratieraamwerk. Het vervolg van deze sectie introduceert elk aspect in meer detail, te beginnen bij het registratie geraamte.

0.2.1 Oppervlakregistratie methodologie

Een algoritme om twee of meerdere oppervlakken te registreren bestaat uit vier hoofdcomponenten: een similariteitsmaat, een transformatiemodel, een doelfunctie en een optimalisatie. We noemen het bewegende oppervlak het oppervlak dat getransformeerd wordt. Het doeloppervlak is het oppervlak waarnaar het bewegende oppervlak getransformeerd wordt. Vooraleer een registratie kan plaatsvinden, moet in eerste instantie een degelijke voorstelling voor het bewegend en doeloppervlak gekozen worden. Een maat van overeenkomst, die gebaseerd is op de gekozen voorstellingen, schat de correctheid van een gegeven transformatie door het getransformeerde bewegend oppervlak te vergelijken met het doeloppervlak. De transformatie zelf is gemodelleerd door en/of gelimiteerd tot een specifiek transformatiemodel, afhankelijk van de voorkennis betreffende het registratieprobleem. Vervolgens worden de similariteitsmaat en het transformatiemodel gecombineerd in een doelfunctie. Tot slot gids de optimalisatie het bewegende oppervlak naar de positie waar de overeenkomst tussen beide oppervlakken maximaal is binnen de spreiding van mogelijke transformaties door de gegeven doelfunctie te optimaliseren naar het globale optimum. Voorafgaande aan de registratie is vaak een initialisatie vereist om de beiden oppervlakken binnen hetzelfde coördinatensysteem te brengen, zodat de optimalisatie kan convergeren naar het correcte globale optimum i.p.v een lokaal optimum.

Oppervlakvoorstelling

Het bewegende oppervlak wordt eenvoudigweg voorgesteld als een verzameling van 3D punten. Voor het doeloppervlak gebruiken we een variationele impliciete oppervlakvoorstelling (VIS), welke de aaneenschakeling is van variationele impliciete functie (VIF) nuldoorgangen. Een VIF wordt gecreëerd gebruikmakend van een variationele verspreide data interpolatietechniek van voorwaarde puntlocaties en functiewaarden. Variationele impliciete functies worden extensief gebruikt doorheen de thesis (niet alleen als oppervlakvoorstelling). Een VIS voorstelling codeert voor elk 3D punt in de ruimte de benaderde euclidische afstand tot en loodrecht op het oppervlak, nul op het oppervlak en positief/negatief voor punten buiten/binnen het oppervlak. Een evaluatie van een punt op het bewegende oppervlak in de VIS geeft de negatieve/positieve afstand tot het doeloppervlak in de richting van de VIS gradiënt wat een interessante inherente voorstellingseigenschap is voor registratiedoeleinden. Wanneer alle punten van het bewegende oppervlak geëvalueerd worden in de VIS van het doeloppervlak wordt er een impliciete notie van overblijvende afstand en verbeteringsrichting gegeven in de punten, zodat de transformatie van het bewegende oppervlak naargelang kan aangepast worden. Met andere woorden, de 3D informatie van het doeloppervlak wordt uitgebreid naar 4D informatie in termen van iso-afstandsoppervlakken tot het doeloppervlak, zodat de vacuüm ruimte tussen de twee oppervlakken opgevuld wordt met informatie betreffende het doeloppervlak.

VIS voorstellingen werden origineel voorgesteld voor gesloten oppervlakken (oppervlakken zonder rand, bvb de schedel in figuur 0.7), omdat het moeilijk is om een notie van de rand van open oppervlakken (partiële en volledige gezichten in de figuren 0.5 en 0.6) te behouden binnen een impliciete oppervlakvoorstelling. Open oppervlakken worden buiten hun rand geëxtrapoleerd en deze extrapolaties kunnen correct en gewenst zijn in de buurt van de rand, maar de correctheid reduceert vrij snel verder weg van de rand. Om dit nadeel van impliciete voorstellingen voor open oppervlakken weg te werken gebruiken we dezelfde gespreide data interpolatietechnieken voor het opstellen van een interne-externe of randfunctie voor open oppervlakken zodat de foutieve extrapolaties gedetecteerd en genegeerd kunnen worden.

Transformatiemodel

Verschillende registratieproblemen vergen verschillende types van transformaties. Het transformatiemodel bepaalt hoe het bewegende oppervlak kan bewegen naar het doeloppervlak. De keuze van een specifiek transformatiemodel reflecteert onze voorafgaande kennis betreffende het registratieprobleem. Afhankelijk van het registratieprobleem en de aanwezigheid van ruis in het doeloppervlak zijn niet alle transformaties mogelijk of realistisch zodat bepaalde voorwaarden geïmponeerd moeten worden opdat de transformatie zich zou gedragen volgens onze voorkennis. Deze voorwaarden worden typisch geïntroduceerd m.b.v een transformatieregularisatie die de ruimte van mogelijke transformaties limiteert.

Maat van overeenkomst

Een similariteitsmaat of maat van overeenkomst geeft de mogelijkheid om de overeenkomst tussen de twee oppervlakken automatisch te evalueren door te definiëren hoe het bewegende oppervlak ideaal gepositioneerd is t.o.v het doeloppervlak. Similariteit kan uitgedrukt worden m.b.v verschillende soorten oppervlakte informatie (vorm of geometrie, buiging, kleur, ...) en is afhankelijk van de gebruikte voorstelling voor het doeloppervlak. Wij poneren een algemene geometriegebaseerde interoppervlak afstandssimilariteit, gemeten in de punten van het bewegende oppervlak. De maat van overeenkomst impliceert dat bewegende oppervlak punten gepositioneerd moeten worden zodat deze punten op een vooraf bepaalde afstand tot het doeloppervlak komen te liggen.

Doelfunctie

Algemeen wordt de constructie van de doelfunctie niet beschouwd als een aparte component voor registratiemethoden. Wij geloven echter dat speciale zorg in acht genomen moet worden wanneer een doelfunctie opgesteld wordt om de robuustheid tegen ruis en onverklaarbaarheden te verhogen. Het nut van een doelfunctie is om de maat van overeenkomst en het transformatiemodel te combineren zodat het optimum van de doelfunctie overeenkomt met een verzameling transformatieparameters die de similariteitsmaat optimaliseren terwijl

de transformatieregularisatie in rekening gebracht wordt. Bijkomend moet de doelfunctie robuust zijn tegen ruis en onverklaarbaarheden in het doeloppervlak die de maat van overeenkomst evaluaties beïnvloeden.

Wij stellen voor om een statistisch maximum a-posteriori (MAP) afleiding van de doelfunctie te hanteren om de similariteitsmaat en het transformatiemodel te combineren. Een dergelijke MAP aanpak heeft al vaak bewezen een nuttig theoretisch raamwerk te zijn, bevoorraad met een krachtige verzameling van goed ontwikkelde mathematische technieken en goed ondersteund volgens een duidelijke filosofische onderbouw. Bovendien zijn alle assumpties klaar en duidelijk geformuleerd. Robuustheid tegen kleine fouten wordt eenvoudig verkregen met zo'n raamwerk door te postuleren dat de similariteitsmaat metingen of evaluaties in de bewegende oppervlakpunten gegenereerd worden door een probabilistisch gemodelleerd verklaarbaarheidproces. Robuustheid tegen grote fouten wordt verkregen door te postuleren dat de similariteitsmaat metingen gegenereerd worden door een statistisch verklaarbaarheidproces of een on-verklaarbaarheidproces die samen een volledig genereringsproces vormen. Hierdoor wordt er een latente variabele geïntroduceerd die de hoeveelheid geloof codeert dat een zekere meting verklaarbaar is of niet. De modellering en de bepaling van de latente variabele is afhankelijk van de voorafgaande kennis omtrent onverklaarbaarheden.

Het finale resultaat is een beter onderbouwde doelfunctie die quasi zelfregelend is in termen van robuustheid tegen kleine en grote fouten. Bovendien kunnen we equivalente M-schatters uit de robuuste statistiek voor het volledige genereringsproces bepalen. Hierdoor kunnen we de invloed van onverklaarbaarheden analyseren en kunnen we gebruik maken van robuuste optimalisatie technieken die gekend zijn binnen het domein van de robuuste statistiek.

Optimalisatie

De optimalisatie definieert hoe de transformatie parameters in de doelfunctie aangepast moeten worden zodat de waarde van de doelfunctie evaluatie daalt overeenkomend met een verbetering in de maat van overeenkomst rekening houdend met de transformatie beperkingen. Optimalisatie is een brede discipline in de wiskunde waar een hele cohorte van methoden voorgesteld zijn met verschillende complexiteit.

Om een doelfunctie, zoals de onze, die afgeleid is van een volledig genereringsproces te optimaliseren is er een gezamenlijke parameter schatting vereist. Buiten de transformatieparameters moeten de latente variabele of parameters ook geschat worden. Deze gezamenlijke parameter optimalisatie kan opgelost worden door afwisselend de latente variabele en de transformatieparameters te verbeteren terwijl tijdens beide optimalisaties de andere parameters ongewijzigd blijven. Dit leidt tot een iteratief dubbele stap verbeteringsroutine die gelijkaardig is aan het welbekende verwachting-maximalisatie (EM) algoritme.

Initialisatie

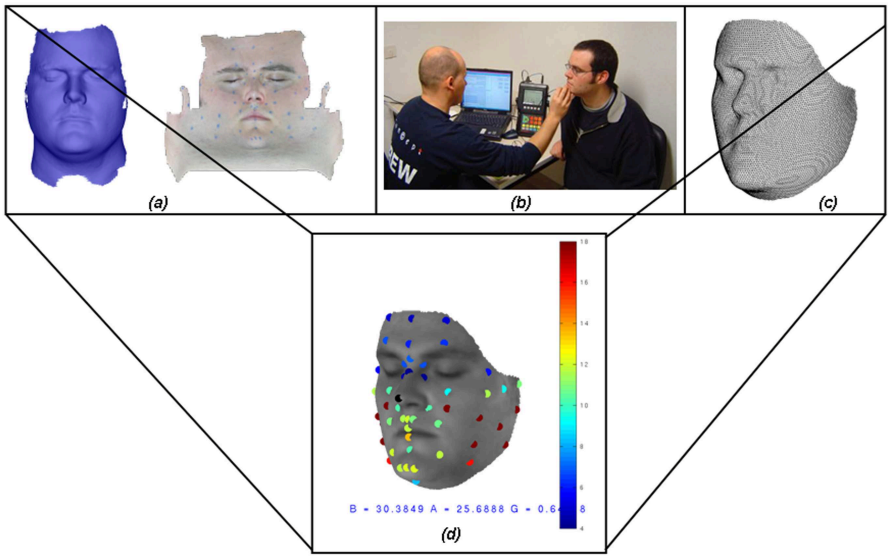
Het voorgestelde registratie algoritme behoort tot de klasse van registratietechnieken waarbij een doelfunctie, die de registratie kwaliteit reflecteert, iteratief geoptimaliseerd wordt. Bijgevolg kunnen deze methode zeer accurate oplossingen genereren en daardoor worden deze methoden ook vaak fijne registratiemethoden genoemd. Echter door hun iteratieve karakter moeten zij geinitialiseerd worden met een schatting van de onbekende transformatie parameters. Voorafgaande aan de registratie moeten het bewegende oppervlak en het doeloppervlak in de buurt van elkaar gebracht worden door een herpositionering van het bewegende oppervlak m.b.v een rigide of affiene transformatie in het coördinatensysteem van het doeloppervlak. Een transformatie schatting voor een registratie wordt typisch een ruwe registratie of initialisatie genoemd.

0.2.2 Databank acquisitie en registratie

Om het gecombineerde statistische gezichtsmodel te creëren, hebben we een databank met gezichtsvoorbeelden die bestaan uit geregistreerde, grijswaarde gekleurde, 3D gezichtsoppervlakken gekoppeld met zachte weefsel diktemetingen (zie figuur 0.4). Om deze databank voorbeelden te verkrijgen hebben we de nodige hardware en software verzameld en ontwikkeld voor het meten en registreren van 3D gezichtsoppervlakken en het meten van zachte weefsel diktes.

Zachte weefsel diktemeting

Een mobiel semi-automatisch ultrageluidsgolf gebaseerd systeem is ontworpen zodat we relatief snel diktemetingen in 52 anatomische merkpunten in het gezicht van levende personen kunnen uitvoeren op een niet invasieve manier. De merkpunten zijn gedefinieerd op basis van voorgaande zachte weefsel diktestudies. Het systeem bestaat uit een compacte, mobiele en digitale ultrageluidsgolf scanner (Epoch 4B, Panametrics Inc., Waltham, USA), een MySQL databank en een zelf ontworpen visueel interactief programma. De Epoch 4B heeft twee grote voordelen vergeleken met traditionele klinische geluidsgolf apparatuur. Ten eerste is de mobiliteit van het toestel optimaal zodat een zo groot mogelijke en diverse populatie opgemeten kan worden. Ten tweede is de diameter van het meetkristal relatief klein zodat merkpunten op moeilijke plaatsen toch meetbaar blijven. De MySQL databank zorgt enerzijds voor de opslag van de meetresultaten en de persoonsgegevens en anderzijds voor de opslag van verschillende Epoch 4B instellingen die afhankelijk zijn van het type persoon en de merkpunten. Deze laatste maken het mogelijk om automatisch de instellingen van de Epoch 4B te wijzigen in functie van het type persoon en het merkpunt. Het interactieve programma controleert de bi-directionele datatransfer tussen de databank en de Epoch 4B zodat automatische dikte berekeningen op basis van het geluidsgolf signaal, automatische bewaring van



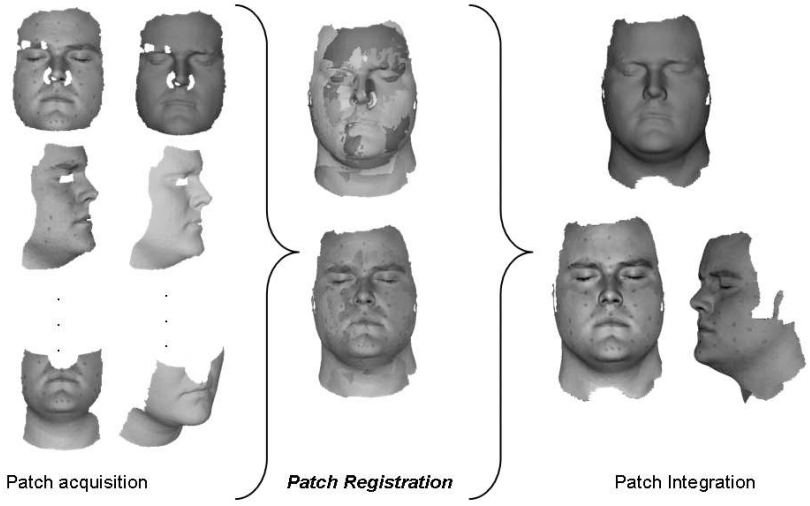
Figuur 0.4: Databank acquisitie en registratie. (a) gezichtsoppervlak meting, (b) zachte weefsel diktemeting, (c) gezichtsoppervlak registratie, (d) Finaal resultaat met de diktemetingen gevisualiseerd m.b.v een kleurencode in de 52 merkpunten.

de resultaten en automatische aanpassing van de instellingen verkregen worden. Het programma zelf wordt bestuurd door een draadloze computermuis met 3 functies. Hierdoor is de uiteindelijke opmeting van een persoon gereduceerd tot maximum 30 minuten waarbij elk van de 52 merkpunten drie maal gemeten wordt om foutieve metingen te voorkomen. De meetprocedure van zachte weefsel diktes wordt geanalyseerd op herhaalbaarheid door bepaalde personen meerdere keren te meten en op accuraatheid door de ultrageluidsgolf metingen te vergelijken met CT gebaseerde diktemetingen.

Gezichtsoppervlak meting

Het gezichtsoppervlak wordt gemeten m.b.v een mobiel 3D fotografisch systeem (ShapeCam, Eyetronics, Belgium). Dit is een actief 3D verwervingsysteem dat een regulier rooster projecteert op het gezicht en waarbij gelijktijdig een digitale foto genomen wordt vanuit een ander kijkstandpunt. Door middel van roosterlijn extracties en triangulaties is het vervolgens mogelijk om 3D informatie van het gezicht te berekenen als een dichte verzameling van verbonden 3D punten. De ShapeCam capteert buiten 3D informatie ook gezichtskleur informatie wat gebruikt wordt voor het extraheren van de 3D merkpunt locaties op het gezicht. Voor het nemen van de 3D foto's worden de merkpunten op het gezicht van de vrijwilliger aangeduid m.b.v een blauw ooglijn potlood. De coördinaten van deze blauwe punten zijn achteraf in de digitale foto's te

extraheren door eenvoudige beeldverwerkingalgoritmen. Vervolgens wordt de volledige kleurinformatie van het gezicht omgezet in grijswaarden. Een groot voordeel van de Shapecam is zijn mobiliteit zoals de Epoch 4B, zodat het eenvoudiger wordt om een grote en diverse populatie op te meten. Bovendien is het perfect mogelijk om gezichten in rechtstaande positie te fotograferen. Een groot nadeel van de ShapeCam, wat algemeen is voor gelijkaardige 3D aquisitie systemen, is de beperkte kijkhoek van de digitale camera (CANON D60) resulterend in een partiële acquisitie van het volledige gezicht per 3D foto opname. De reden hiervoor is dat het systeem enkel en alleen 3D informatie kan berekenen van datgene dat zichtbaar is vanuit een zeker kijkstandpunt. Daarom, moeten meerdere partiële opnames gemaakt worden vanuit onbekende kijkstandpunten (door de camera mobiliteit) om een volledige 3D gezichtsopname te verwerven.



Figuur 0.5: Tijdens een eerste fase moeten de gezichtsstukken accuraat gepositioneerd worden in een gezamenlijk coördinatenstelsel: de registratiefase. Ten tweede moeten de verschillende gezichtsstukken geïntegreerd worden tot een enkelvoudige entiteit: de integratiefase.

Het combineren van meerdere partiële gezichtsopnames of gezichtstukken tot een enkelvoudig volledig gezicht vergt twee hoofdfases die geïllustreerd zijn in figuur 0.5. Tijdens een eerste fase moeten de gezichtsstukken accuraat gepositioneerd worden in een gezamenlijk coördinatenstelsel: de registratiefase. Ten tweede moeten de verschillende gezichtsstukken geïntegreerd worden tot een enkelvoudige entiteit: de integratiefase. Bovendien moeten zowel de registratie als de integratiefase robuust zijn tegen ruis en onverklaarbaarheden afkomstig van het opnameproces en door de partiële overlap tussen verschillende gezichtstukken. Om een volledig automatische assemblage te hebben op basis van partiële opnames zonder enige voorkennis betreffende de camera

standpunten moeten meerdere registratietaken opgelost worden. Eerst moet een initiële positionering bepaald worden d.m.v een ruwe initialisatie die vervolgens verfijnd moet worden d.m.v een fijne registratie. Het eerste is een globaal pose probleem of initialisatie waar geen voorkennis is gegeven betreffende de relatieve oriëntaties en positionering van de gezichtstukken. Het tweede of de pose verfijning veronderstelt een gegeven initialisatie. Wanneer men meer dan twee partiële opnames heeft, moeten zowel de paarsgewijze ruwe initialisaties als de fijne registraties geconverteerd worden naar een meerdere standpuntregistratie of positionering. Tot slot, na de registratiefase, moeten de gepositioneerde stukken geïntegreerd worden. In plaats van voor elke deeltaak een bestaand algoritme te selecteren en te gebruiken zou het interessant zijn om een gezamenlijke oppervlakvoorstelling te gebruiken doorheen de volledige assemblage procedure zelfs tot en met de integratiefase.

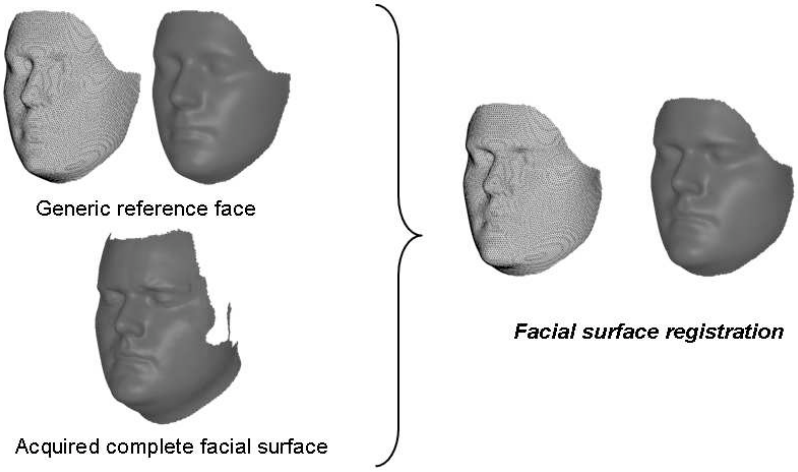
Om een volledig automatische, accurate en robuuste registratie en integratiemethode van partiële opnames, zonder enige voorkennis van het opnameproces, te verkrijgen gebruiken we een 4-staps algoritme gebaseerd op een gemeenschappelijk geheugenvriendelijke VIS voorstelling van de open partiële oppervlakken. Deze impliciete voorstelling bevat interessante eigenschappen om de verschillende registratie en integratie problemen te behandelen. Bovendien, kunnen zacht verlopende eisen gehanteerd worden om met ruis in de partiële oppervlakken om te gaan. In een eerste stap worden robuuste maar minder accurate ruwe initialisaties tussen elk paar van partiële opnames bepaald gebaseerd op punt correspondenties, die verkregen worden door de vergelijking van georiënteerde punthandtekeningen berekend op basis van de variationele impliciete oppervlakvoorstelling. Een pose classificatie procedure selecteert de beste rigide transformatie schatting uit mogelijke transformaties die de overeenkomstige punten naar elkaar toe transformeren terwijl het hun oriëntatie doet samenvallen. Lokale 2D geometrie histogrammen berekend op basis van de VIS voorstelling worden gebruikt als punthandtekeningen, VISS-beelden genaamd, welke een verbeterde variant zijn van de traditionele spinbeelden. In een tweede stap wordt een selectie gemaakt van de paarsgewijze ruwe initialisaties om deze te converteren naar een meerder standpunt ruwe initialisatie d.m.v een minimaal omspannende boom algoritme (MST). De paarsgewijze transformatiekost wordt uitgedrukt in functie van het aantal transformatie ondersteunende puntcorrespondenties. Het resultaat van het MST algoritme is een niet-redundante minimale verzameling van relatieve paarsgewijze transformaties om alle gezichtstukken in een gezamenlijk coördinatenstelsel te initialiseren. Tijdens de derde stap worden de verzameling van initiële transformaties verder verfijnd door het uitvoeren van paarsgewijze fijne registraties, die minder robuust zijn in termen van initialisatie maar die wel accurate resultaten kunnen genereren. De fijne registraties bestaan erin om een afstand gebaseerde doelfunctie, o.b.v de continue VIS voorstelling, iteratief te optimaliseren volgens een dalende gradiënt methode. Om het probleem van Partiële overlap tijdens de registratie op te lossen wordt er een nieuwe deterministische onverklaarbaarheid functie geconstrueerd die de latente variabele modelleert en die onafhankelijk is van de relatieve positionering of afstandhis-

togrammen tussen twee partiële gezichtsstukken. Zowel de optimalisatie als de onverklaarbaarheids behandeling zijn gedefinieerd en gecombineerd in het voorgestelde robuuste en statistisch gebaseerde registratieraamwerk. Hierbij is het bewegende oppervlak een zeker gezichtsstuk en het doeloppervlak een ander gezichtsstuk terwijl we een rigide transformatiemodel gebruiken. Tot slot, in de laatste stap, wordt de accumulatie van paarsgewijze registratie fouten weg gewerkt door een simultane meerder standspunt fijne registratie van alle partiële opnames met een tussentijdse integratie uit te voeren. Deze fijne registratie van een gezichtsstuk met een tussentijdse integratie van al de andere stukken is zeer gelijkaardig aan de paarsgewijze fijne registratie. Dezelfde gemeenschappelijke VIS oppervlakvoorstelling wordt gebruikt doorheen alle stappen van de procedure tot en met de integratiefase zodat de registratie en de integratie gecombineerd kunnen worden in de vierde en laatste stap. Buiten vormintegratie wordt er ook een additionele kleurintegratie uitgevoerd. Zowel vorm als kleurintegratie zijn gebaseerd op een gewogen combinatie van de VIS voorstellingen en variationele kleurfuncties, respectievelijk, van de individuele partiële opnames. De gewichtsfuncties, welke lokaal een maat van oppervlakcorrectheid weergeven, zijn ook verkregen op basis van de gespreide data interpolatie technieken (variationele impliciete functies). Het finale resultaat is een volledig, enkelvoudig, gezichtsoppervlak met kleurinformatie geassembleerd vanuit meerdere partiële gezichtsopnames. De complete vier stap procedure wordt uitgevoerd op realistische en gesimuleerde data om de nodige validatie evaluaties te verkrijgen en te analyseren.

Gezichtsoppervlak registratie

Om een statistische analyse te verkrijgen betreffende de gezichtsvorm informatie in de databank moeten alle gezichten niet-rigide geregistreerd worden. Een enkelvoudig volledig gezichtsoppervlak bestaat uit een wolk van 3D punten die onderling verbonden zijn. Geregistreerde gezichtsoppervlakken hebben hetzelfde aantal 3D punten met dezelfde connectiviteit alleen beschrijven de 3D punten een andere geometrische vorm. Om geregistreerde gezichten in de databank te verkrijgen transformeren we de 3D punten met zekere connectiviteit van een referentiegezicht naar alle gezichtsoppervlakken in de databank door te zoeken naar overeenkomstige punten. Het resultaat is dat elk gezicht in de databank evenveel punten heeft met dezelfde connectiviteit zodat voor een bepaald punt op het referentiegezicht we de corresponderende punten op alle andere gezichten kennen. Omdat de 52 gezichtsoppervlak merkpunten een deelverzameling zijn van de volledige gezichtspunten kennen we bijgevolg ook de corresponderende merkpunten op alle gezichten.

De gezichtsoppervlak registratiemethode, die voorgesteld wordt in figuur 0.6, moet robuust zijn tegen ruis en onverklaarbaarheden door partiële overlap of ontbrekende 3D informatie. De robuustheid tegen ontbrekende data is omdat de 3D acquisitiescanner niet in staat is om 3D informatie te berekenen in regio's bedekt met haar zodat finale volledige gezichtsoppervlakken soms gaten en ontbrekende data vertonen. Een voorbeeld is zichtbaar in figuur 0.6,



Figuur 0.6: De 3D punten met zekere connectiviteit van een referentiegezicht worden getransformeerd naar alle gezichtsoppervlakken in de databank door te zoeken naar overeenkomstige punten. Links boven, het referentiegezicht gevisualiseerd als roostervoorstelling en een computergrafische visualisatie. Links onder een gezichtsoppervlak uit de databank en rechts het geregistreerde oppervlak uit de databank als roostervoorstelling en computergrafische visualisatie

waar er geen 3D informatie voor handen is op de plaats van de bakkebaarden. De robuustheid tegen ruis, is niet zozeer door fouten in de doeloppervlakken (zijnde volledige gezichtsoppervlakken in de databank), omdat deze weggewerkt werden tijdens de partiële oppervlak integratiefase, maar eerder door het niet-rigide karakter van het registratieprobleem. Een verwezenlijking van overeenkomstige punten tijdens de registratie is typisch foutief, vooral in het begin van de registratie omdat de vormen van het referentiegezicht, zijnde het bewegende oppervlak, en de doeloppervlakken dan erg verschillend kunnen zijn.

Het complete gezichtsregistratie probleem wordt opgelost d.m.v het voorgestelde registratieraamwerk. Hierbij maken we gebruik van een niet-rigide TPS gebaseerd transformatiemodel en een maat van overeenkomst o.b.v de continue VIS doeloppervlak voorstelling. Onverklaarbaarheden door ontbrekende data vormen geen probleem door gebruik te maken van een latente variabele modellering die gelijkaardig is aan diegene van de fijne partiële opname registraties. Door het niet-rigide karakter van het registratieprobleem maken we gebruik van een deterministische annihilering optimalisatie. Resultaten worden bekeken in termen van nauwkeurigheid en consistentheid. De nauwkeurigheid validatie gaat na of de getransformeerde 3D punten van het referentiegezicht de vorm van het doelgezicht accuraat beschrijven. De consistentheid validatie gaat na of corresponderende punten op verschillende gezichten effectief corre-

sponderend of consistent aangeduid zijn.

0.2.3 Gezichtsmodellering en reconstructie

Gebaseerd op de verkregen databank bouwen we een gezichtsmodel dat gebruikt wordt voor gezichtsreconstructies. De meest abstracte formulering van onze reconstructie procedure is: Wat is het meest waarschijnlijke gezicht, volgens een statistisch gezichtsmodel, gegeven de fout bevattende schedeldata. Het gezichtsmodel wordt verkregen door een statistische modellering van de data in de databank resulterend in een statistisch gezichtsmodel. Het meest waarschijnlijke gezicht gegeven de schedeldata wordt verkregen door een probabilistische of statistische gebaseerde reconstructie aanpak. Een nodige en voldoende schedeloppervlak voorstelling en een transformatiemodel worden bepaald om een statistisch geformuleerde gezichtsmodel naar schedelregistratie uit te voeren die als dusdanig het meest waarschijnlijke gezicht genereert.

Statistische gezichtsmodellering

Om foutieve maskergerelateerde invloeden in de gezichtsreconstructies te vermijden en om het waarschijnlijke karakter van de reconstructies te verhogen stellen wij een statistisch gezichtmodel voor reconstructies voor. Wij modelleren de gecombineerde populatie-afhankelijke variatie en co-variatie van volledig gezichtsoppervlak vorm met grijswaarde kleur, 52 anatomische merkpunten met bijhorende diktemetingen en eigenschapwaarden (BMI, leeftijd, geslacht) informatie, berekend o.b.v een uitgebreide databank verkregen m.b.v een fotometrische 3D scanner en ultrageluidsgolf technologie.

Het modelmasker of referentiegezicht kan geclassificeerd worden als zijnde een volledig gezichtsoppervlak waaruit een enkelvoudige meest waarschijnlijke reconstructie bekomen wordt o.b.v geanalyseerde of statistische kennis van meerdere gezichten. De statistische modellering wordt bekomen m.b.v principale component analyse zo dat het gezichtsmodel beschouwd kan worden als een elastisch masker met elastische halve bollen aan de binnenkant. De elasticiteit van het masker wordt bepaald door de statistisch toegestane gecorreleerde variatie van verschillende complete gezichtsoppervlakken en zachte weefsel diktemetingen. Door de model transformatieparameters te veranderen binnen de statistische grenzen wordt het masker op een gezichtsspecifieke manier vervormd en wordt de gezichtsplausibiliteit gegarandeerd. Zodoende wordt het masker op de schedel geplaatst totdat de halve bollen het schedeloppervlak raken.

De creatie van meer subpopulatie gerichte gezichtsmodellen volgens gegeven schedeleigenschappen, verkregen uit een antropologisch onderzoek, wordt bekomen door de expliciete modellering en eliminatie van variaties afkomstig van verschillen in individuele eigenschapwaarden in de databank. Het resultaat is een genormaliseerd gezichtsmodel in functie van de gegeven schedeleigenschappen. De incorporatie van grijswaarde informatie genereert een meer levendige reconstructie in functie van de vorm en dikte informatie. Wij geloven dat

het gebruik van grijswaarde i.p.v kleurinformatie minder gevaarlijk is om een herkenning in de foute richting te duwen. Kleurinformatie is altijd moeilijk te schatten op basis van de schedeldata en moet gelimiteerd worden tot het geven van een idee (idee van lippen, wenkbrauwen, ...) zonder al te veel details.

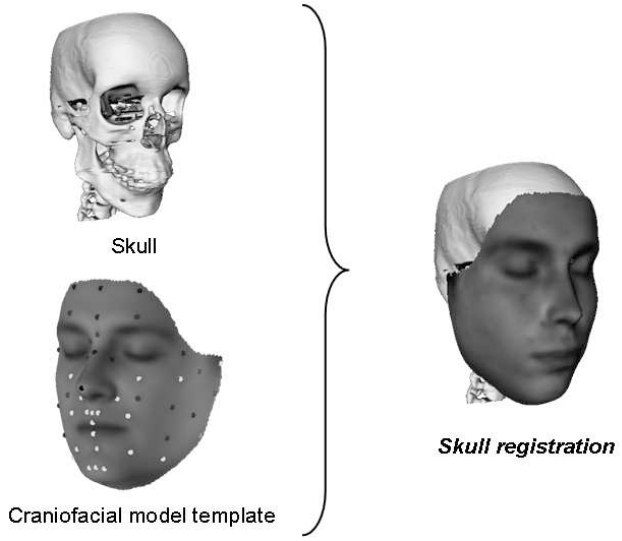
Het grootste verschil van ons gezichtsmodel met meer traditionele gezichtsmodellen is de incorporatie van co-variatie kennis tussen verschillende gezichtsinformatie delen. Door achtereenvolgens co-variatie informatie te elimineren bekomen we een traditioneel gezichtsmodel dat gebruik maakt van een enkelvoudig referentiegezicht in combinatie met een generisch (niet gezichtsspecifiek) transformatiemodel.

De kwaliteit van het statistische gezichtsmodel aangaande de vorminformatie is afhankelijk van de punt correspondenties verkregen door de volledige gezichtsoppervlak registratieprocedure. De meest belangrijke voorwaarde voor model opbouw is de consistente aanduiding van de punt correspondenties op de gezichtsoppervlakken in de databank. Dit is echter niet eenvoudig te valideren door een gebrek aan controlegegevens. Een voldoende consistentie evaluatie kan wel verkregen worden door na te gaan of de gecombineerde kennis betreffende correspondenties in een trainingsverzameling van gezichten in staat is om de correspondenties van een ongezien gezicht te beschrijven. Gebruikmakend van het statistische vormmodel kunnen we een dergelijke test uitvoeren. Het statistische model bevat namelijk de gecombineerde kennis van correspondenties afkomstig van meerdere gezichten en kan gebruikt worden in het registratieraamwerk om modelgebaseerde correspondentie te bepalen op een ongezien gezichtsoppervlak. Deze modelgebaseerde correspondenties kunnen dan vergeleken worden met de eerder bekomen TPS gebaseerde correspondenties om na te gaan of beide correspondenties voor hetzelfde gezicht consistent zijn. M.a.w, de modelgebaseerde correspondenties zijn de controlegegevens. Verder wordt er nagegaan of de statistische modellering voldoende is om enerzijds gezichten in de databank gebruikt voor de modelopbouw te beschrijven en anderzijds ongeziene gezichten niet gebruikt voor de modelopbouw te beschrijven.

Statistische gezichtsreconstructie

Het zoeken van de ontbrekende gezichtsenveloppe horende bij een onbekende schedel is het doel van de gezichtsreconstructie en wordt opgelost door een model naar schedel of kortweg schedelregistratie (zie figuur 0.7). Vooraleer een schedelregistratie kan plaatsvinden, moet een virtuele kopie van de schedel gemaakt worden voor een computergebaseerde reconstructietechniek. Vervolgens moet een gezichtsmodel en een schedeloppervlak voorstelling gekozen worden zodat een maat van overeenkomst kan opgesteld worden. Men mag echter niet vergeten dat een virtuele kopie nooit een exacte kopie is en dat bijkomende ruis en onverklaarbaarheden geïntroduceerd worden tijdens de schedelvoorstelling opbouw. Hierdoor moet de schedelregistratie robuust zijn tegen ruis en onverklaarbaarheden. Tot slot, is een uitvoerige validatie vereist opdat de ontwikkelde methode bruikbaar wordt in de praktijk tijdens foren-

sische onderzoeken en gerechtelijke veroordelingen.



Figuur 0.7: Het modelmasker wordt getransformeerd naar de schedel om een gezichtsreconstructie te verkrijgen tijdens een schedelregistratie.

De virtuele kopie van de schedel wordt bekomen d.m.v een medische CT scanner. Onze primaire keuze van gezichtsmodel is het statistische model van de vorige sectie. Een automatische gezichtsreconstructie wordt verkregen op basis van een impliciete schedelvoorstelling, welke onze primaire voorstellingskeuze is. Een variationeel impliciet oppervlak van de schedel wordt gecreëerd op basis van de gespreide data interpolatietechnieken. Deze impliciete voorstelling codeert voor elk 3D punt de benaderde euclidische afstand tot en loodrecht op het schedeloppervlak, nul op het oppervlak en negatief/positief voor punten binnen/buiten het oppervlak. De similariteitsmaat nodig voor de registratie is gebaseerd op de vergelijking van de VIS evaluaties met de zachte weefsel diktes in de 52 merkpunten van het modelmasker. Het doel van de schedelregistratie bestaat er dan in om de 52 merkpunten en diktemetingen te herpositioneren en te veranderen zodat de diktemetingen gelijk worden aan de VIS evaluaties in de merkpunten. Echter is de relatie tussen VIS evaluaties en ultrageluidsgolf diktemetingen niet 100% correct zodat de similariteitsmaat metingen beïnvloed worden door ruis en onverklaarbaarheden. Een tweede, meer traditionele, schedelvoorstelling keuze is gebaseerd op 52 manueel aangeduide merkpunten en hun normaal op het schedeloppervlak in overeenstemming met de 52 gezicht merkpunten op het gezichtsmodel masker. De similariteitsmaat tussen het model en de schedelvoorstelling is gebaseerd op de meetprocedure van zachte weefsel diktes met ultrageluidsgolf technologie. De optimalisatie tijdens de schedelregistratie komt er dan op neer om de schedeldiktes te meten terwijl het gezicht gereconstrueerd wordt. Echter zijn de aangeduide schedel

merkpunten niet foutloos en afhankelijk van de expertise van de antropoloog die de punten aanduidt zodat de similariteitsmaat beïnvloed wordt door ruis en onverklaarbaarheden.

Om het meest waarschijnlijke gezicht te vinden volgens de gezichtsmodel keuze gegeven de gekozen schedelvoorstelling maken we gebruik van het voorgestelde registratieraamwerk. Het doeloppervlak is het schedeloppervlak. De keuze van gezichtsmodel levert het bewegende oppervlak en een bijhorend transformatiemodel. De maat van overeenkomst is gedefinieerd o.b.v de gezicht-schedel relaterende informatie in het gezichtsmodel en de keuze van schedelvoorstelling en wordt beïnvloed door ruis en onverklaarbaarheden. De statistisch geformuleerde schedelregistratie procedure construeert en optimaliseert een robuuste doelfunctie die het transformatiemodel en de maat van overeenkomst combineert. De latente variabele voor onverklaarbaarheden wordt probabilistisch gemodelleerd m.b.v een random variabele en een statistisch onverklaarbaarheid proces.

Een eerste validatie van de voorgestelde gezichtsreconstructie aanpak is gebaseerd op een klinische CT databank van 12 patiënten en simuleert realistische reconstructie scenario's. Elk gezicht in de databank wordt om de beurt gebruikt als testcase dat de nodige controlegegevens bevat. De resulterende gezichtsreconstructies op basis van de gegeven schedel worden dan vergeleken met het bijhorende gezichtsoppervlak zijnde de controlegegevens. Op basis van deze databank worden beide schedelvoorstelling keuzes vergeleken. Verder worden ook verschillende gezichtsmodel keuzes, met verschillende hoeveelheid co-variate kennis, vergeleken. Hierdoor kan nagegaan worden of de incorporatie van co-variatie tussen verschillende gezichtsinformatie delen al dan niet een positieve bijdrage levert tot de kwaliteit van de berekende reconstructies. Tot slot, gaan we ook na of de robuuste formulering van het registratieraamwerk nodig is.

Een tweede validatie wordt bekomen o.b.v de databank voor gezichtsmodel opbouw om meer gecontroleerde ruis en onverklaarbaarheden simulaties te verkrijgen. Elk gezicht in de databank wordt om de beurt uit de databank genomen en als testgezicht gebruikt, terwijl de andere gezichten gebruikt worden om het gezichtsmodel op te bouwen. Door het gebrek aan volledige schedeloppervlakken in de modeldatabank kunnen we de impliciete schedelvoorstelling hier niet gebruiken. In plaats daarvan gebruiken we een beter gecontroleerde merkpunt gebaseerde schedelvoorstelling in vergelijking met de patiëntendatabank. Om de invloed van kleine fouten of ruis te analyseren voegen we incrementeel kleine fouten op de foutloze schedel merkpunten toe. Om de invloed van onverklaarbaarheden te analyseren voegen we grote fouten toe aan een incrementeel percentage van licht foutieve merkpunten.

0.3 Resultaten

De praktische relevantie en de algemeenheid van het robuuste statistisch gebaseerde oppervlakken registratieraamwerk wordt verkregen door het toe te

passen en te testen voor de verschillende registratieproblemen. Wij passen het raamwerk toe op de registratieproblemen die opgelost moeten worden in de volledige gezichtsreconstructie procedure.

De testen betreffende het voorgestelde vier-staps assemblagealgoritme voor volledig gezichtsoppervlak metingen zijn voldoende accuraat en robuust om de methode in de praktijk te gebruiken. De testen worden uitgevoerd op praktisch relevante data en gesimuleerde data. De praktische data toont aan dat de methode inderdaad robuust is tegen slechte partiële opnames. Foutloze gesimuleerde data illustreert dat de methode accuraat is en foutbevattende gesimuleerde data toont aan dat de methode kan werken met ruis in de partiële oppervlak opnames. De VISS-beelden die gebruikt worden als punthandtekeningen leveren een beter ondersteunde en robuustere initile pose schatting om de fijne registraties te initialiseren vergeleken met de traditionele spin-beelden. De nieuw gedefinieerde onverklaarbaarheid functie maakt het mogelijk om eenvoudiger met partiële overlap te werken vergeleken met traditionele alternatieven zodat de finale fijne registratieresultaten meer dan submillimeter zijn en bijgevolg zeer accuraat. Door gebruik te maken van de geheugen efficiënte VIS voorstellingen worden een aantal tekortkomingen van huidige integratie methoden opgelost.

De accuraatheids - en consistentietesten betreffende de niet-rigide registratie van volledige gezichtsoppervlakken, m.b.v een TPS transformatiemodel, geven bevredigende resultaten. De geregistreerde gezichtsoppervlakken beschrijven de geometrie van het oorspronkelijke oppervlak met een voldoende nauwkeurigheid. Door gebruik te maken van een statistisch model in het registratieraamwerk wordt de consistentie nagegaan van de TPS gebaseerde puntcorrespondenties en hieruit blijkt dat de correspondentie kwaliteit bruikbaar is in de praktijk. M.a.w de bepaalde puntcorrespondenties zijn kwalitatief voldoende om een nuttig en correct statistisch gezichtsmodel te creëren. Het gezichtsmodel zelf heeft daardoor een goede modelleringkracht om gezichten die gebruikt tijdens de modelopbouw en gezichten die niet gebruikt worden tijdens de modelopbouw te beschrijven of te modelleren met voldoende nauwkeurigheid.

De reconstructie procedure wordt geëvalueerd op basis van de patiënten en modeldatabank. Op basis van de patiënten databank worden beide schedelvoorstelling keuzes vergeleken en zien we dat de impliciete voorstelling een waardige keuze is om te gebruiken in de praktijk. Bijkomend verkrijgen we met de impliciete keuze ook een volledig automatische gezichtsreconstructie procedure. Bovendien blijkt het ook dat de merkpunt aanduiding voor de traditionele schedelvoorstelling keuze niet eenvoudig is en dat de resultaten hierdoor sterk beïnvloed worden. Verder worden ook verschillende gezichtsmodel keuzes, met verschillende hoeveelheid co-variate kennis, vergeleken. Hieruit zien we dat de voorgestelde statistisch gezichtsmodel keuze superieur is vergeleken met de meer traditionele gezichtsmodel keuzes die gebruik maken van een generisch transformatiemodel. Tot slot gaan we ook na of de robuuste formulering in het registratieraamwerk nodig is op basis van deze patiënten databank. Dit toont aan dat er inderdaad rekening gehouden moet worden

met ruis en onverklaarbaarheden in de schedelvoorstellingen en dat het voorgestelde registratieraamwerk krachtig genoeg is om er mee om te gaan. De testen o.b.v de modeldatabank bevestigen wederom de voorgaande conclusies omtrent de gezichtsmodel keuzes. Bijkomend zien we dat mits een goede paramater keuze het statistisch gebaseerde robuust schedelregistratie raamwerk zeer nuttig is tegen verschillende niveaus van ruis en onverklaarbaarheden.

0.4 Besluit

Forensische gezichtsreconstructie bestaat erin om een schatting te genereren van de gezichtsenveloppe horende bij een onbekende schedel. Alle reconstructietechnieken zijn gebaseerd o.b.v van de veronderstelde relaties tussen de schedel en de zachte weefsels. Manuele technieken worden reeds uitvoerig gebruikt in de praktijk maar zijn echter tijdrovend en subjectief. Computergebaseerde methoden zijn interessant om het reconstructieproces te versnellen terwijl de subjectiviteit ingeruild wordt voor objectiviteit. Al de bestaande methoden hebben een overeenkomstige manier van werken waarbij een gezichtsmodel, in functie van schedeleigenschappen, naar een virtuele kopie van de schedel geregistreerd wordt op basis van een maat van overeenkomst. Deze laatste wordt gedefinieerd in functie van een gekozen schedelvoorstelling en de gezicht-schedel relaterende informatie die vervat zit in het gezichtsmodel.

Wij stelden een complete statistische aanpak voor om gezichtsreconstructies te genereren. De filosofie achter de aanpak bestond erin om het meest waarschijnlijke gezicht te zoeken volgens een gezichtsverdeling gegeven de fout bevattende schedeldata. Een statistisch gezichtsmodel werd gebouwd en gebruikt om de verkeerde modelmasker invloeden te elimineren en om de gezichtswaardigheid van de reconstructies te verbeteren. Hiervoor werd de gecombineerde variatie en co-variatie gebruikt van gezichtsvorm en zachte weefsel diktemetingen vervat in een uitgebreide databank. Het meest waarschijnlijke gezicht werd verkregen door een statistisch geformuleerde schedelregistratie te hanteren. Bijkomend werd hierdoor de robuustheid tegen kleine en grote fouten verhoogd.

De hele gezichtsreconstructie procedure vergde een aantal deeltaken waarin er verschillende soorten oppervlak registratieproblemen opgelost moesten worden. Een oppervlak registratie bestond erin om de geometrische relatie tussen twee of meerdere oppervlakken te bepalen. Hiervoor werd een algemeen statistisch gebaseerd robuust registratieraamwerk gedefinieerd en ontworpen. Robuustheid tegen ruis en onverklaarbaarheden werd verkregen door een volledig genereringsproces bestaande uit een verklaarbaarheidproces en onverklaarbaarheidproces te definiëren. Hierdoor werd een extra latente variabele geintroduceerd die het geloof reflecteerde dat een similariteitsmaat meting al dan niet verklaarbaar was. De manier van modellering voor de latente variabele hing af van de voorkennis betreffende het voorkomen van onverklaarbaarheden. Het raamwerk werd gehanteerd om de verschillende registratieproblemen op te lossen en de nodige validatie testen werden uitgevoerd. Uit deze testen bleek

dat het registratieraamwerk krachtig en nuttig is om de nodige robuustheid en accuraatheid te verwerven voor uiteenlopend oppervlakregistratie doeleinden.

Toekomstige richtingen betreffende de registratie methodologie bestaan erin om additionele oppervlakinformatie (bvb buiging en kleur), buiten vorm, mee te incorporeren in de maat van overeenkomst. Verder onderzoek over de gezichtsmodellering bestaat erin om de ultrageluidsgolf diktemetingen eventueel te vervangen door VIS evaluatie op basis van een databank bestaande uit volledige schedeloppervlakken en bijhorende gezichtsoppervlakken. Hierdoor is de relatie in de similariteitsmaat op basis van een impliciete schedelvoorstelling tijdens registratie minder foutief zodat er meer registratiebewijs is om de reconstructie te verbeteren. Bijkomend kunnen meer punten met diktemetingen op het gezichtsoppervlak van het modelmasker gebruikt worden vergeleken met de gelimiteerde 52 merkpunten, vermits de VIS diktemetingen in de databank voor modelopbouw dan automatisch uitgevoerd kunnen worden.

Forensische gezichtsreconstructie balanceert tussen de werelden van wetenschap en kunst. Dankzij de inzet van vele mensen in de wereld wordt de lijn langzaam maar zeker verlegd richting wetenschap. Zo kan de praktische relevantie van gezichtsreconstructies tijdens politionele onderzoeken en gerechtelijke veroordelingen stijgen ten voordele van het slachtoffer. Ik ben overtuigd dat de waarde van gezichtsreconstructies sterk kan toenemen wanneer de reconstructietechnieken gecombineerd worden met gezichtsidentificatie technieken, wat momenteel een actief onderzoeksdomein op zich is.

List of acronyms and symbols

Abbreviations

1D	:	one-dimensional
2D	:	two-dimensional
2.5D	:	two and a half-dimensional
3D	:	three-dimensional
4D	:	four-dimensional
BMI	:	body mass index
CC	:	cross correlation
CFR	:	craniofacial reconstruction
CG	:	contaminated gaussian (estimator)
CM	:	cranio-metric
CP	:	closest point
CT	:	Computer tomography
DNA	:	desoxyribonucleic acid
EM	:	expectation maximization
E-step	:	expectation-step
GT	:	ground truth
ICP	:	iterative closest point
KL	:	Kullback-Leibler
LM	:	landmark
LMSQ	:	least median of squares
LSQ	:	least-squares
LV	:	latent variable
LV-step	:	latent variable-step

M-step	:	maximization-step
M-estimator	:	maximum likelihood estimator
MAD	:	median of absolute deviations
MAP	:	maximum a posteriori
MC	:	marching cubes
MD	:	mahalanobis distance
MF	:	mean-field
MI	:	mutual information
MIC	:	medical image computing
ML	:	maximum likelihood
mm	:	millimetre
MRI	:	magnetic resonance imaging
MST	:	minimum spanning tree
MT	:	marching triangles
PCA	:	principal component analysis
PDF	:	probability density function
Q	:	quadratic (estimator)
RBF	:	radial basis function
RGB	:	red-green-blue
RMSE	:	root mean squared error
RPM	:	robust point matching
sDT	:	signed distance transform
SSD	:	sum of squared differences
SVM	:	support vector machine
TPS	:	thin plate spline
TQ	:	truncated quadratic (estimator)
US	:	ultrasound
USA	:	united states of America
VBF	:	variational boundary function
VIF	:	variational implicit function
VIS	:	variational implicit surface
VISS-image	:	variational implicit surface spin-image
VMF	:	variational mapping function
VTF	:	variational texture function
VWF	:	variational weighting function

X-rays	:	röntgen rays

Symbols

general aspects

$\mathbb{E}$	:	3D/4D Euclidean space
$\mathbb{R}$	:	real numbers
$f(\mathbf{r})$	:	regular continuous function
$\mathbf{r} = (x, y, z)$	:	3D point with x, y and z coordinates
$\mathbf{n} = (n_x, n_y, n_z)$	:	normal
$\partial f(\mathbf{r})/\partial x$	:	partial derivative of f to x
∇f	:	gradient of the function f, vector of partial derivatives
$\nabla_{conj} f$	:	Conjugate gradient of the function f
$A, \mathbf{a}$	:	matrix, vector
$A^{\mathrm{T}}, \mathbf{a}^{\mathrm{T}}$	:	matrix, vector transpose
$\widehat{\mathbf{a}}$	:	argmax or argmin solution
I	:	identity matrix
MD	:	mahalanobis distance

variational functions

$\mathfrak{f}(\mathbf{r})$	:	VIF, variational implicit function
$\mathbf{r}_i, h_i$	:	variational function constraint points (centers) and values
$v(\mathbf{r})$	:	a low degree polynomial
ϕ	:	RBF, radial bases function, TPS in our case
Φ	:	RBF kernel, TPS kernel in our case
λ_i, Λ	:	RBF coefficient and coefficient matrix
$\|\mathfrak{L}()\|^2$	:	regularization functional
ν	:	user amount of regularization
$\mathfrak{d}(\mathbf{r})$	:	VIS, variational implicit surface
$\mathfrak{b}(\mathbf{r})$	:	VBF, variational boundary function
$\mathfrak{w}(\mathbf{r})$	:	VWF, variational weighting function
$\mathfrak{m}(\mathbf{r})$	:	VMF, variational mapping function

$\mathfrak{t}(\mathbf{r})$	:	VTF, variational texture function

surface registration framework

S, S^f, S^t	:	surface, floating surface, target surface
N, N^f, N^t	:	number of surface, floating surface and target surface points
∂S	:	surface boundary or border
$\exists S$	:	surface interior
^{t}	:	(superscript) target element instance
^{f}	:	(superscript) floating surface element
$\mathfrak{T}(S, \boldsymbol{\theta})$	:	surface transformation
$\mathfrak{T}(\mathbf{r}, \boldsymbol{\theta})$	:	3D point transformation
$\boldsymbol{\theta}$	:	transformation model parameters
$\widehat{\boldsymbol{\theta}}$	:	MAP estimated transformation model parameters
$\nu\|\mathfrak{L}(\boldsymbol{\theta})\|^2$	:	transformation regularization, log model-likelihood
$p(x), p(x\|y)$	:	probability, conditional probability
$\mathfrak{s}_j$	:	individual point similarity
$\mathfrak{D}(S, \mathbf{r})$	:	current distance from a 3D point $\mathbf{r}$ to the surface S
d_j	:	predefined distance
ϵ	:	additive error
σ	:	(error) standard deviation
$\Theta = \{\boldsymbol{\theta}, \sigma\}$	:	complete set of ML parameters (noise + transformation model parameters)
$p(S\|\Theta), p(\boldsymbol{\theta})$	:	complete data-likelihood and model-likelihood
$\mathfrak{E}$	:	MAP objective function
$\mathfrak{S}$	:	complete log data-likelihood
$\rho(x, \gamma)$	:	M-estimator with x the residual error and γ a scale parameter
$\rho(x, \alpha, \beta)$	:	M-estimator originating from an outlier-process with α and β parameters
Ψ	:	M-estimator outlier influence function
$z, \mathbf{z}$	:	outlier latent variable and map
$p(\mathbf{z})$	:	prior outlier probability
$h(x, \alpha(z), \beta(z))$	:	complete-process (outlier + inlier-process) function
δ	:	uniform distribution probability

$p^b(x)$	:	Bernoulli probability distribution
$b_j, 1-b_j$	:	individual point inlier- and outlier-belief
λ	:	prior outlier parameter
κ	:	statistical significance prior outlier parameter
$Q(x\|\widehat{x})$	:	Q function, expectation of the complete log data-likelihood
$E[x\|y]$	:	expectation of x conditioned on y
s	:	gradient or angle (viss-images) step

facial surface acquisition

(u, v)	:	(viss-images)image plane coordinates
γ	:	(viss-images) attenuation parameter
Im	:	(viss-images) image plane
$\mathbf{b}_i$	:	(viss-images) image bin center
$corr(Im_1, Im_2)$	:	normalized correlation coefficient between two images
P	:	number of patches
R	:	rotation matrix
$\mathbf{t}$	:	translation vector
$\omega_x, \omega_y, \omega_z$	:	euler angles
t_x, t_y, t_z	:	x, y and z translations
$H(x)$	:	Heaviside function
S^{int}	:	integrated surface
$D(\mathbf{r}), \mathfrak{D}(\mathbf{r})$	:	Volumetric patch integration, without and with VIFs
$\mathbf{s}$	:	viewing direction
$\mathrm{t_d}, \mathrm{t_w}$	:	distance and confidence thresholds
$\mathfrak{K}$	:	subset of patches influencing a certain point
$\mathfrak{R}, \mathfrak{G}, \mathfrak{B}$	:	integrated red, green and blue functions

facial surface registration

C	:	rigid or affine transformation matrix using homogenous coordinates
$\mathfrak{J}$	:	set of inlier points
r	:	annealing rate
T_σ	:	annealing parameter

facial modeling

$\mathbf{F}, \bar{\mathbf{F}}, \breve{\mathbf{F}}, \tilde{\mathbf{F}}$	:	facial vector, average face, mean and poperty normalized face
$\mathbf{F}^s, \mathbf{F}^g$	:	facial surface and texture information
$\mathbf{F}^{lm}, \mathbf{F}^{td}$	:	facial landmark and tissue depth information
$\mathbf{F}^p$	:	facial property information
g, d, b, a, s	:	gray values, thickness values, BMI, age and gender values
X	:	covariance matrix
$\mathbf{U}_k, U$	:	eigenvector and eigenvector matrix
$M(\mathbf{c})$	:	Model with parameters $\mathbf{c}$
σ_k^2	:	eigenvalues
$\mathbf{v}, V$	:	property basis vector and matrix
$\triangle x$	:	BMI, age or gender difference

facial reconstruction

$\mathfrak{H}$	:	Hessian matrix
F_{est}	:	estimated facial outlook
$\mathbf{v}$	:	measuring direction
ϵ_d, σ_d	:	error and standard deviations on measured distance (cranio-metric representation)
ϵ_a, σ_a	:	error and standard deviations on measuring angle (cranio-metric representation)
$rmse_l$	:	global rmse for a single reconstruction
$rmse$	:	overall averaged global rmse over all reconstructions
$rmse_i$	:	overall local averaged rmse over all reconstructions
$\mathbf{F}^{GT}$	:	ground truth facial information
$\mathbf{F}^R$	:	reconstruction facial information
d_A	:	facial similarity defined as the angle between facial parameterizations
$\breve{\mathbf{g}}$	:	scaled model parametrization
CM	:	cranio-metric skull representation
VIS	:	implicit skull representation

Contents

Chapter 1

Introduction

1.1 Problem statement

Why CFR? Sunday morning, 7:00 AM, the sky is clear and the sun is rising. Mr. Anonymous is jogging along the canal, when suddenly he discovers the skeletal remains of a person in the flowering shrub next to the water. He immediately calls the authorities with his cellular phone and 20 minutes later the Police arrives at the scene. Before any other investigation can be performed the primary question to be answered is: Who is this person? The answer is obtained by forensic identification techniques which are mostly based on comparisons of ante- and post-mortem data, such as medical files, dental records, X-rays or DNA. This whole procedure becomes less evident when dealing with skeletonized human remains, where any link with a possible identity is missing. In these circumstances, a *craniofacial reconstruction (CFR)* might help the investigation out of the impasse. The goal of craniofacial reconstruction is to recreate an estimate of the face of an individual at the time of death. Different 2D and 3D manual or computer-aided facial reconstruction techniques have been developed for this purpose and all are based on the assumed relationship between the soft tissue envelope and the underlying skull substrate.

From manual to computerized CFR Several 3D manual methods for facial reconstruction have been developed and are currently used in practice. These reconstructions consist of physically modeling a face on a skull replica (the target skull) with clay or plasticine. However, manual reconstruction methods require a lot of anatomical and artistic modeling expertise and are as a result highly subjective. Furthermore, these reconstructions take a lot of time, and, hence, are often limited to a single reconstruction. Computer-based methods, on the other hand, are consistent and objective. Moreover, since these methods can be executed in a short time, multiple reconstructions from the same skull using different modeling assumptions (older, thicker, ...) can be obtained. The development of software for computerized facial approximations

of an individual would be of benefit to various law enforcement agencies, by allowing faster, easier and more efficient generation of multiple representations of an individual.

CFR Framework Computerized reconstruction techniques all share a common work-flow in which a craniofacial model, linking the skull structures to the soft tissue facial envelope, is warped onto a virtual copy of the skull specimen. The craniofacial model consists of a facial mask combined with a geometric transformation model. The registration of the model towards the skull is the process of finding and applying the geometrical transformation between the craniofacial model and the skull. The registration is driven by a similarity measure expressing the goodness of fit of the craniofacial model to the skull. Both the craniofacial transformation model and the similarity measure are combined into an objective function, to ensure a proper optimization. During optimization a set of transformation model parameters is searched for that optimize the objective function, increasing the similarity model fit, while assuring a well-behaved transformation. Once these parameters are known, the complete facial model template can finally be deformed towards the skull specimen to recreate an estimate of the face.

What is lacking? Current computerized reconstruction techniques are limited, though, in the model used for reconstructing the complete facial outlook. First, either a generic face template or a specific best look-alike template, based on skull similarities or properties, is chosen. Subsequently, the skin surface associated to the target skull is estimated by transforming the model template, based on a generic, "smooth"deformation (e.g. TPS based). Multiple reconstructions based on different values of BMI, age and gender are obtained by choosing a different starting facial template. Two major shortcomings are apparent using such a model template in combination with a generic deformation. First, the reconstruction can be incorrectly biased by the choice of the template. Indeed, when using a subject-specific best look-alike template based on similarity in ancestry, gender and age, unwanted facial features of the template remain visible in the final reconstruction. Using a generic face template, on the other hand, results in too smooth and unspecific a reconstruction. Secondly, the generic deformations applied are not face-specific, they are just "smooth". No problem arises when the differences between the model and target skull are small. However, when these differences are relatively large, the required deformation will be more pronounced, which can result in unrealistic, caricature-like or implausible facial reconstructions.

What is required? The appearance of biased and face-unspecific reconstructions can be reduced by the use of a face-specific model *database* built over ideally, a large, representative set of attribute-labeled (race, gender, age, BMI) facial masks and associated face-to-skull information. The facial masks are preferably acquired in an upright position. The information relating the

skull substrate to the soft tissue facial envelope can be spatially *sparse*, based on a limited number of anatomical landmarks at which the soft tissue thickness is measured by ultrasound probing, e.g. *Dense* point sets, on the other hand, can be obtained, e.g., from CT-scans, although still mostly acquired in a supine position and over a patient population. In any case, the type of skull-to-skin link information will also determine the choice of an appropriate *surface representation* for the unknown skull specimen. An important issue concerning the registration of the craniofacial model towards the skull is the presence of *small errors or noise and gross errors or outliers* within the skull representation. Besides the fact that a virtual copy is never an exact but a noisy copy of the skull specimen, additional errors are typically introduced during skull representation build-up. Until now, to the best of our knowledge, no robust (in terms of noise and outliers) skull registration framework has been defined or used in current computerized craniofacial reconstruction techniques. A last issue concerning computerized reconstruction techniques is *validation*. From an engineering point of view, a developed method can work fine, but is useless in practice unless properly validated. A proper validation framework will hopefully substantiate on a scientific basis the added value of reconstruction methods during crime-scene investigations.

1.2 Scope of the thesis

Robust Statistical CFR In this thesis we present as a first major contribution, a computerized (semi-)automated craniofacial reconstruction method that is both consistent and objective. It builds on a statistical craniofacial model of the interrelationship between facial shape and soft-tissue thicknesses as measured on a predefined set of landmark points on the skull. The model can be considered as an elastic mask with elastic hemi-spherical dowels on the inside of the mask at the landmark locations, which is subsequently fitted to the skull such that the virtual dowels touch the skull. The elastic deformation of the transformation model is based on the statistical likelihood as learned from a database of nearly 400 subjects. This requires several subtasks.

1. The acquisition of 3D skin surfaces and tissue-depth information in an upright position, measured over a sufficiently large and diverse population which are subsequently stored in a database together with the subject's age, Body Mass Index (BMI), gender and ancestry information.

2. The establishment of inter-subject dense surface correspondences for further statistical modeling.

3. Statistical modeling of the population-dependent variation and covariation of skin surface shape and tissue depth, fine-tuned towards a given set of skull properties (ancestry, age,...).

4. Fitting of this statistical model to the individual craniofacial skeleton.

5. Validation of the complete craniofacial reconstruction approach.

Robust, statistical, implicit surface registration framework The whole procedure (1-5) involves a number of surface registration tasks. Surface registration refers to the establishment of the geometrical relationship between two or more surfaces. It is a key enabling technology to solve a wide range of 3D computer-based modeling and reconstruction problems. As a second major contribution of this thesis, we present a robust, statistical, surface registration framework that optimizes an inter-surface distance measure over the parameters that define the geometric transformation. We use a memory efficient, continuous, smooth and analytical, implicit surface representation, which has particularly interesting properties for registration purposes. Robustness against noise and outliers of the surface models during registration is obtained using a Maximum a Posteriori (MAP) formulation including a statistical inlier- and outlier-process model. This framework is applied to solve different registration tasks necessary in the complete craniofacial reconstruction procedure or pipeline: First in subtask (1), complete facial surfaces are assembled from partial surface patches using a rigid transformation model. Second in subtask (2), surfaces of different subjects are matched to each other using a non-rigid Thin Plate Spline based deformation model. In subtask (3), the observed inter-individual deformations are statistically modeled via Principal Component Analysis, assuming a multivariate normal distribution. This model is matched in a first instance to the facial surfaces of the different subjects in the database using a transformation model based on the learned statistical deformations to analyze the quality of the Thin Plate Spline based registrations in the previous subtask. Finally in subtask (4), this facial model (including the soft-tissue thicknesses) is aligned with the (virtual) skull specimen using again the transformation model based on the learned statistical deformations.

The aim of this thesis is to solve every subtask within the craniofacial reconstruction procedure, using the proposed surface registration framework. The remainder of this section will introduce every aspect in more detail, starting with the developed surface registration framework.

1.2.1 Surface registration methodology

An algorithm to register and transform two or more surfaces consists of four main components: a similarity measure, a transformation model, an objective function and an optimizer. We call the *floating* surface the surface that is to be transformed. The *target* surface is the goal surface, i.e. the surface the floating surface is transformed to. In a first instance a proper *surface representation* for both the floating and target surface is to be chosen. The *similarity measure*, which is dependent on the chosen surface representations, estimates the correctness of a given transformation by comparing the transformed floating surface with the target surface. The transformation itself is modeled by and/or restricted to a specific *transformation model*, describing the type of geometric transformations allowed (rigid, affine, non-rigid, ...). The similarity

measure and the transformation model are combined within an appropriate error, energy or *objective function*. Finally, the *optimizer* estimates the transformation parameters that optimally map the floating surface onto the target surface. Often, prior to surface registration, an initialization is required to appriximately align the floating and target surface ensuring the optimizer to converge to the correct global optimum of the objective function.

Surface representation

The choice of surface representation has a direct influence on the accuracy and efficiency for surface registration. In this thesis, the floating surface is simply represented as a set of 3D points. The target surface, on the other hand, is represented as a variational implicit surface (VIS). Here, the surface is defined as the locus of zeros of the associated variational implicit function. A variational implicit function (VIF) is created from constraint points using a variational scattered data interpolation approach and is extensively used throughout the thesis (not only for surface representations). The VIS representation encodes for every 3D point an approximate shortest Euclidean distance to the target surface, zero on the target surface, positive outside and negative inside. Evaluation of a floating surface point into the target VIF gives the negative/positive distance to the target surface in the direction of the VIS gradient in that point, which is a very interesting property for registration purposes. Hence, when the points of the floating surface are evaluated into the VIS of the target surface, a notion of remaining distance and updating direction at these points to the target surface is implicitly given. We also extend the concept of VIS, originally proposed for closed surfaces, to open surface as well using the very same machinery of implicit function representations.

Transformation model

Different registration problems require different types of transformations. The transformation model stipulates how the floating surface can be transformed to the target surface. The choice of using a particular type of transformation model reflects our prior knowledge on the registration problem. Depending on the registration problem and the presence of noise within the target surface not all transformations are feasible or realistic so that certain mapping constraints are imposed to ensure that the transformation behaves according to our prior knowledge. These constraints impose a transformation regularization, restricting the space of possible transformations. We will discuss generic (face-unspecific) and statistical (face-specific) transformation models.

Similarity Measure

A similarity measure provides an automated evaluation of the match between the two surfaces to register by defining how the floating surface is to be positioned ideally with regard to the target surface. Similarity can be expressed using different kinds of surface information (e.g. geometry, texture, curvature)

and is dependent on the target surface representation. We propose a general *geometric-based* inter-surface distance similarity, measured in the floating surface points, implying that the floating surface points are to be repositioned or transformed such that the points lie at predefined distances to the target surface.

Objective function

Although generally not considered being a separate component for registration, we believe that special care is to be taken when constructing an objective function to increase robustness against noise and outliers within the similarity measure. The purpose of such an objective function is to combine the similarity measure and the transformation model such that the optimum (minimum or maximum) of the function generates a set of transformation parameters while improving the similarity measure and taking the transformation regularization into account. Furthermore, the objective function must be robust against noise or small errors and outliers or gross errors in the target surface data influencing the similarity measure.

We propose to use a statistical Maximum a Posteriori (MAP) formulation of the objective function to combine the similarity measures and the transformation regularization. MAP approaches have proven to be very useful theoretical frameworks, equipped with a set of powerful and well developed mathematical techniques. Furthermore, all the assumptions are explicit. Robustness against noise is easily obtained by stating that the similarity measures, evaluated at the floating surface points, are generated by a statistically modeled noisy inlier-process. Robustness against outliers is obtained by stating that the same similarity measures are generated by either a probabilistic inlier- or an outlier-process, both constituting a complete-process. In order to do so, an outlier latent variable is introduced encoding the presence of an inlier/outlier at a floating surface point. The modeling and determination of the outlier latent variable is dependent on the prior knowledge concerning the outlier generating process.

The final result is a better underpinned objective function with quasi self-regulating robustness against noise and outliers. Furthermore, we define equivalent, closely related, M-estimators from Robust Statistics for the complete-process formulation. Doing so, outlier influences in the objective function can be analyzed and well known optimization techniques from Robust Statistics can be applied.

Optimization

The optimization method defines how the parameters within the objective function are adjusted or updated to improve the objective function evaluation and therefore improving the matching quality while taking into account the transformation model regularization. Optimization is a broad discipline in

mathematics where a lot of methods have been proposed with varying complexity.

To optimize the proposed objective function derived from a complete-process modeling a joint parameter estimation is to be performed. Indeed, besides the transformation parameters, also outlier latent variables are to be estimated as well. This joint optimization can be solved by alternating between a latent variable optimization and a transformation parameter optimization while during either optimization or update the other update or estimate is kept fixed.

Initialization

The proposed surface registration method belongs to a class of registration algorithms in which an objective function, reflecting the quality of registration, is iteratively optimized. Consequently these methods can generate accurate results and are referred to as *fine registration* methods. However, due to their iterative nature they need to be initialized with a proper estimate of the unknown pose parameters. Prior to registration, the floating surface is to be brought into the vicinity of the target surface, by repositioning the floating surface, using a rigid or affine transformation, into the coordinate system of the target surface. A pose estimation prior to fine registration is typically referred to as *crude registration* or *initialization*.

1.2.2 Database acquisition and registration

In order to build the required statistical craniofacial model, we acquire a database of three-dimensional facial entries or samples (figure 1.1 (d)), consisting of registered 3D gray-textured skin surfaces coupled with soft-tissue depths measured at 52 anatomical landmarks on the skin. For acquisition of the 3D facial entries within the database, we assembled acquisition hardware and software to capture and register 3D skin surfaces and to measure tissue depths.

Thickness acquisition

For measuring the *soft tissue depths* a mobile semi-automated ultrasound system (figure 1.1(b)) enabling in vivo, fast and non-destructive measurements at the 52 anatomical landmarks (10 midline plus 2 times 21 bilateral) is developed. These landmarks are defined in and are measured according to [1, 2]. The system is composed of a compact and lightweight mobile digital ultrasound "A-mode"scanner (Epoch 4B with a 10MHz 0.6mm diameter transducer, Panametrics Inc., Waltham, USA), a database (MySQL) and a self-designed interface program.

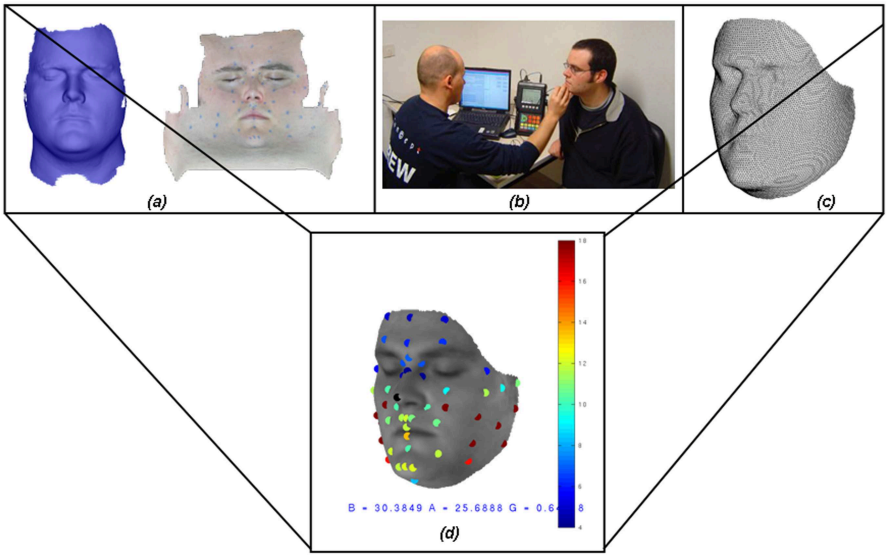


Figure 1.1: Database acquisition and registration. (a) skin surface acquisition, (b) soft-tissue depth measurement, (c) facial surface registration, (d) final result with the tissue depths visualized by means of a color code in the 52 face landmarks.

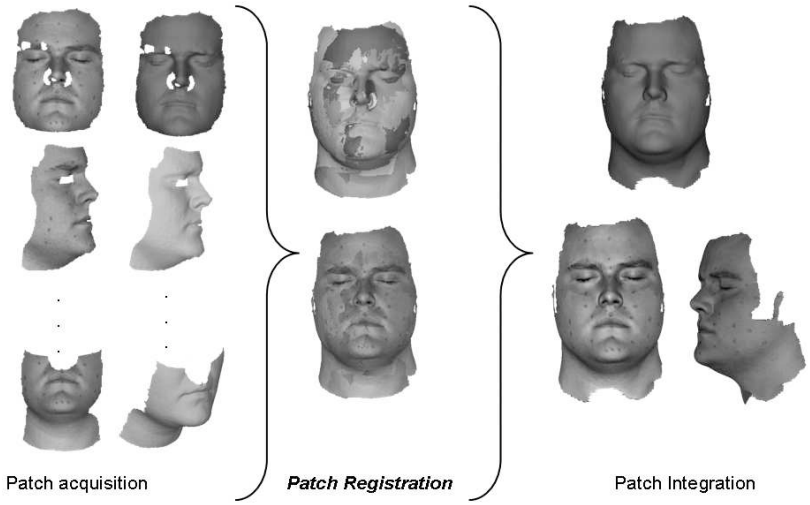


Figure 1.2: Patch registration involves the accurate alignment of multiple partial surface acquisitions into the same coordinate system. Subsequently the patches are integrated into one complete surface entity

Facial surface acquisition

The *skin surface* (figure 1.1(a)) is measured by a mobile 3D photographic device (ShapeCam, Eyetronics (www.eyetronics.com)). This is an active 3D capturing device, which projects a regular grid on the face. At the same time a digital camera (CANON D60) takes an image of the face from a different point of view such that by triangulation the 3D shape of the face can be retrieved as a dense set of connected 3D points. The ShapeCam also captures *texture information*, which is used to determine the 3D location of the landmarks on the skin surface reconstruction. A disadvantage of the ShapeCam, which is common for similar 3D acquisition devices, is the limited viewing-angle of the camera, resulting in a partial surface acquisition of the complete facial surface. The reason is that the acquisition system can only compute 3D information of what can be "seen" from a certain viewpoint. Therefore, a complete three-dimensional facial surface is assembled from several partial surface acquisitions or patches from unknown viewpoints (due to the camera portability).

Combining several patches into a single surface typically involves two main phases, which is shown in figure 1.2. First, the patches need to be aligned accurately into a common coordinate frame, the *registration phase*. Second, the registered patches need to be integrated into a single entity, which is the *integration phase*. Furthermore, both the registration and the integration must be robust against noise and outliers in the patches due to the acquisition process and due to the partial overlap in between different patches. In order to have a fully automatic assembly, without any kind of a priori knowledge of the geometry of the imaging process or user interaction, more than one registration task has to be solved. In chapter 4, we present a fully automatic, robust and accurate method for aligning and integrating partial acquisitions without any prior knowledge of the relative viewpoints of the camera or the geometry of the imaging process. This multi-step registration and integration algorithm is based on memory efficient VIS patch representations and used the surface registration framework proposed in chapter 3.

Facial surface registration

To perform a statistical analysis of 3D shape information all faces in the database need to be *registered* non-rigidly. A single skin surface is represented by a dense set of connected 3D points. Registered skin surfaces share the same amount of 3D points with the same connectivity. For this, the 3D points with known connectivity of a carefully constructed generic reference face are mapped onto the faces in the database (figure 1.1(c)) by finding dense point correspondences. The result is that every facial surface in the database is represented by the same amount of points with the same connectivity, such that for every point on one surface the corresponding point on every other surface in the database is known. Because the 52 face landmarks are a subset of the dense points representing skin surfaces, corresponding inter-subject face landmarks are known as well.

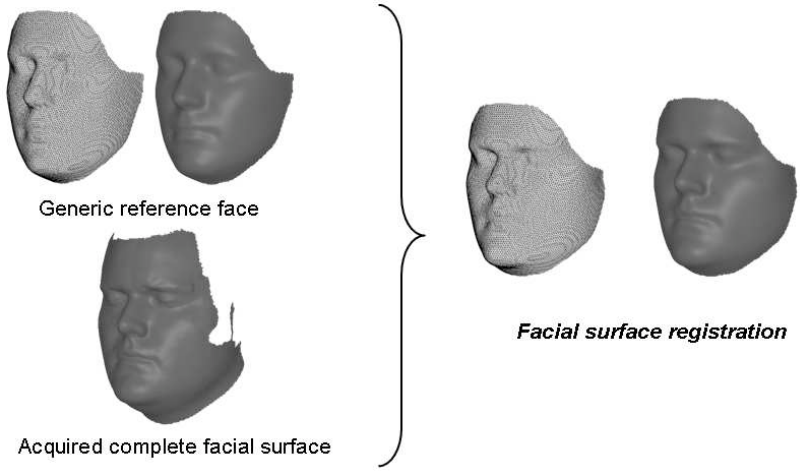


Figure 1.3: The 3D points with known connectivity from a reference face are mapped onto the faces in the database during facial surface registration

The facial surface registration methodology, illustrated in 1.3 must be robust against noise and must be able to cope with outliers due to partial overlap or missing data. The latter is because of the fact that the 3D surface acquisition device is not able to acquire 3D information from hairy or covered regions in the face, generating incomplete 3D surface acquisitions with holes or missing data. An example is visible in figure 1.3, due to the presence of side-whiskers the 3D facial surface contains holes or misses 3D surface information at the location of the whiskers. The former, or noise robustness, is not necessarily due to small errors in the target surface (being complete facial surfaces from the database), because these were smoothed after shape integration, but rather due to the non-rigid nature of the registration. An instance of point correspondences during registration is typically erroneous, especially at the beginning of the registration, because the shapes of the floating and the target surface are then very different.

To tackle the facial surface registration undertaking robustly, we propose to apply the presented robust statistical surface registration framework based on a non-rigid thin-plate-spline (TPS) transformation model and a similarity measure deducted from a VIS target surface representation. Outliers, due to partial overlap or missing data, are dealt with using a deterministic outlier detection function and probabilistic outlier-process, similar to the outlier handling during fine registrations of patches. Because of the non-rigid and complex nature of the registration problem a deterministic annealing optimization is applied.

1.2.3 Craniofacial modeling and reconstruction

Based on the acquired database we construct a craniofacial model used for craniofacial reconstruction. The core most abstract formulation of the presented craniofacial reconstruction approach is: *What is the most plausible or probable face, according to a statistical craniofacial model being a face distribution, given the (erroneous) skull data*. The face distribution is obtained through the *statistical modeling* of the data in the craniofacial database, resulting in a statistical craniofacial model. The most plausible face given the skull data is obtained using a probabilistic or *statistical reconstruction* process. A proper skull representation and craniofacial transformation model are determined to conduct a probabilistic formulated craniofacial model to skull registration, generating the most plausible solution accordingly.

The presented computer-based reconstruction method in this thesis is also a further refinement of the work we have published in [3, 4, 5, 6, 7]. It must be noted that some final fundamental changes have been made within the craniofacial model and skull representation to improve the reconstruction quality and skull representation vs model compatibility. Furthermore, a new and improved probabilistic registration framework, applicable to a wide range of craniofacial models and skull representations, is proposed to increase robustness against noise and outliers in the skull representation.

Statistical craniofacial modeling

In order to eliminate the template-related model bias and to minimize the unrealistic character of the reconstructions caused by large generic model deformations, we propose a flexible statistical craniofacial model for reconstruction. We model the combined population-dependent variance and covariance of complete skin surface shape, 52 anatomical face landmarks with tissue depths, property (age, BMI and gender) values and skin surface gray-value texture information, calculated from an elaborate facial database constructed with a photogrammetric 3D scanner and ultrasound tissue depth measuring technology. Note that in previous published work, thicknesses were set out starting from face landmarks on the skin surface perpendicular to the skin surface, resulting in 52 skull landmarks per face in a database. However this lead to incorrectly determined skull landmarks for every face, because thicknesses are measured perpendicular to the skull surface instead of the face surface. Therefore, thicknesses are not set out anymore, but instead a thickness is considered being an extra coordinate for the 52 face landmarks.

The major difference of the proposed statistical model with existing models for craniofacial reconstruction is the incorporation of covariance knowledge between different facial information parts. Eliminating subsequently covariance between different information parts from the model leads to more traditional craniofacial models working with a single facial template in combination with generic deformations.

Statistical craniofacial reconstruction

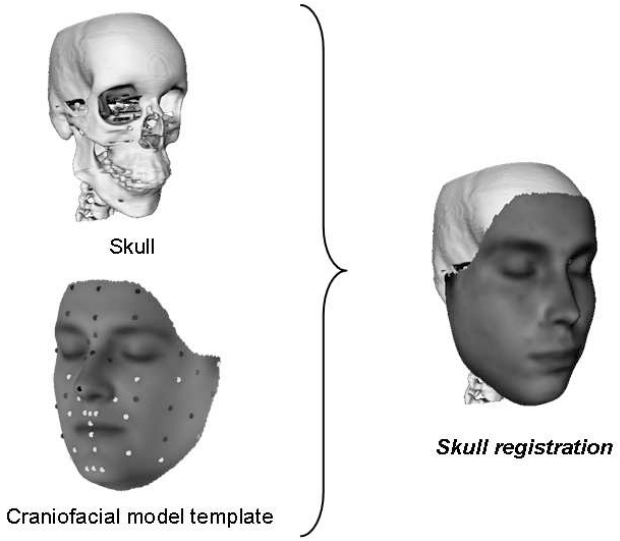


Figure 1.4: The craniofacial model template is deformed or adapted towards a given skull surface during the skull registration

Finding the facial outlook of a given unknown skull is the purpose of a craniofacial reconstruction and is solved during the skull registration illustrated in figure 1.4. In a first instance a virtual copy of the skull is required for a computer-based reconstruction technique. In a second instance a proper choice of craniofacial model, reflecting our prior knowledge on the reconstruction problem, and a skull representation, being compatible with the craniofacial model to establish a proper similarity criterium, are required. However, one has to keep in mind that a virtual copy of the skull is never an exact copy and that additional noise and outliers are incorporated during skull representation build up such that the skull registration must be robust against noise and outliers. Finally, a proper validation framework is required to substantiate on a scientific basis the added value of the reconstruction method during crime-scene investigations.

The virtual copy of the skull substrate is obtained using a CT scanner. Our primary choice of craniofacial model is the statistical model from the previous section. An automatic reconstruction procedure is obtained based on a primary proposed implicit skull representation choice. A variational implicit surface of the skull is created using variational scattered data interpolation techniques. This implicit surface representation encodes for every 3D point an approximated shortest euclidian distance perpendicular to the skull surface, zero on the surface, positive outside and negative inside. The similarity measure needed for registration is based on the comparison of the VIS values, being

signed distance values, with the ultrasound tissue depths in the 52 model template landmarks. The goal of the skull registration is to reposition the 52 face landmarks and to re-estimate the tissue depths such that the VIS values resemble the tissue depths. However, the link between VIS values and tissue depths is never 100% correct such that the similarity measure is corrupted by noise (small errors) and outliers (gross errors). A second, more traditional, skull representation choice, leading to a semi-automatic reconstruction, is based on 52 manually indicated craniometric skull points (corresponding with the 52 model template landmarks) and surface normal information (in contrast with previous work). The similarity criterium between this craniometric skull representation and the model is based on the procedure for measuring the tissue depths in the database with ultrasound measuring technology. Trying to optimize this similarity measure during registration is then similar to measuring the skull while reconstructing the face. However, the manual skull landmark indications are error-prone and dependent on the expertise of the anthropologist conducting the reconstruction, such that the similarity measure is corrupted by noise and outliers.

To find the most plausible face according to our craniofacial model given the skull representation we propose to use the robust statistical surface registration framework. The target surface is the skull surface. The choice of craniofacial model is responsible for the floating surface, being the model mask or template, and the transformation model based on the learned statistical deformations. The similarity measure is defined based on the choice of skull representation and is corrupted by noise and outliers. The statistical formulated skull registration constructs and optimizes a robust objective function combining a transformation model and similarity measure, while being robust against noise and outliers. Outlier modeling is done using a probabilistic random outlier latent variable in combination with the outlier-process model.

A first validation of the proposed craniofacial reconstruction method is based on a database of 12 clinical patient cases and is used to simulate realistic reconstruction scenarios. Each facial sample is removed, in turn, from the patient database and used as a test case. The resulting reconstructed facial skin surfaces can then be compared quantitatively and qualitatively, with the skin surface of the test case representing the ground truth. Based on this database a comparison between the results obtained with the craniometric and the implicit skull representation is made. Furthermore a comparison between the results obtained with different craniofacial models incorporating various amounts of covariance within the same skull registration framework and based on the same skull representations is made to validate the contribution of covariance incorporated in the proposed statistical craniofacial model. Finally, the contribution of the skull registration framework for robust model registration is validated as well based on the clinical patient database.

A second validation is obtained by a leave-one-out cross-validation procedure performed on the database used to build up the model and is used for simulating controlled noisy and outlier skull data points in a craniometric representation. Each facial sample is removed, in turn, from the model database

and used as a test case. The remaining facial entries are used to create the different craniofacial models. Because of the lack of complete skull surface information in the model database, the implicit skull representation cannot be used. Instead we use a better controlled, in terms of noise and outliers compared to the patient database, craniometric skull representation. To examine the robustness against small errors in the skull representation we gradually insert gaussian distributed noise onto the skull-landmark positions. Additionally, to examine the robustness against gross errors we inserted various amounts of outliers within the skull representation.

1.3 Thesis outline

The structure of the thesis consist of three parts and is mainly determined by the craniofacial reconstruction application. The first part is quite general, introduces the CFR application and presents the underlying general methodology. The second and third part comprise the necessary chapters of the complete CFR pipeline, which is schematically illustrated in figure 1.5. The second part in combination with Appendix C concentrate on the facial database acquisition and registration. The third part presents the statistical craniofacial modeling and reconstruction approach.

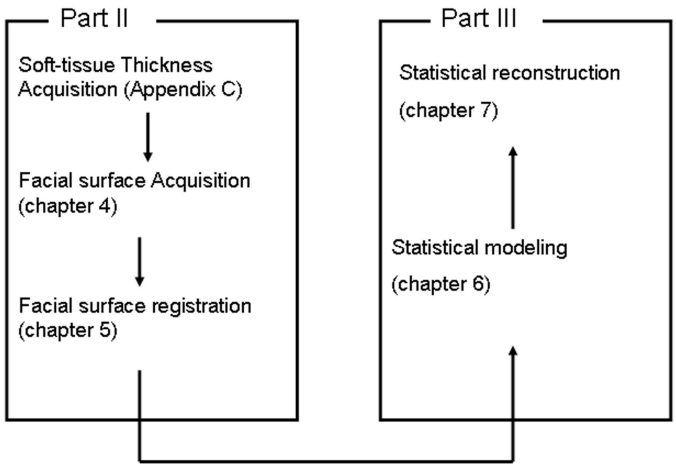


Figure 1.5: Schematic overview of the chapters constituting the CFR pipeline, Database building (Part II) containing registered facial surfaces and tissue-depths, Craniofacial modeling and reconstruction (Part III)

Part I: Application and general methodology

- Chapter 2: Defines and explains a general work-flow for computer-based craniofacial reconstruction techniques while enumerating related work concerning the main application.
- Chapter 3: Explains the robust statistical surface registration framework using implicit function representations.

Part II: Facial database acquisition and registration

- Chapter 4: Concentrates on the facial surface assembling from several partial surface or patch acquisitions.
- Chapter 5: Explains the establishment of dense correspondences on inter-subject facial surfaces needed for statistical modeling.

Part III: Craniofacial modeling and reconstruction

- Chapter 6: Elaborates the statistical modeling of the acquired database in order to obtain a statistical craniofacial model. Furthermore, statistical model manipulation according to a given set of skull property values is explained as well.
- Chapter 7: In this chapter the statistical approach for craniofacial reconstruction is given and applied on different craniofacial model and skull representation choices. Furthermore, an extensive validation concerning the several choices and the skull registration framework is given as well in this chapter.
- Chapter 8: Gives a final conclusion and future work issues.

Appendices

- Appendix A: Two main approaches for robust statistics are presented and applied to the surface registration problem.
- Appendix B: A principled approach to link robust estimators presented in appendix A to a statistical model of how the measurements, including the outliers, were generated is presented.
- Appendix C: Details concerning the ultrasound thickness acquisition system are given in this appendix.
- Appendix D: Gives an elaborate related work overview concerning patch registration and integration.
- Appendix E: A short related work overview concerning the non-rigid facial surface registration for model building is presented.

- Appendix F: Explains how a nonnumeric gender property can be converted into a continuous numeric value in order to incorporate the gender into the statistical modeling.

1.4 Main contributions

The major contributions of this thesis are:

- The definition of a general-work flow for computer-based craniofacial reconstruction techniques combined with a complete overview of existing methods.

- The development of a robust statistical surface registration framework using implicit function representations that optimizes an inter-surface distance measure over the parameters that define the geometric transformation. We use a continuous, smooth and analytical, implicit surface representation, which has particularly interesting properties for registration purposes. Robustness against noise and outliers of the surface models during registration is obtained using a Maximum a Posteriori (MAP) formulation including a statistical inlier- and outlier-process model. This framework is applied to solve different registration tasks necessary in the complete craniofacial reconstruction procedure or pipeline

- The construction of a variational boundary function to retain the interior-exterior concept within implicit representations for open surfaces.

- In order to provide a fully automatic, robust and accurate method for aligning and integrating partial reconstructions without any prior knowledge of the relative viewpoints of the camera or the geometry of the imaging process, we developed a 4-step registration and integration algorithm based on common memory efficient VIS representations of the individual patches.

  - Construction of VISS-images, being point signatures computed based on VIS representations, to obtain an automatic crude pair-wise initialization.
  - Application of the surface registration framework based on VIS representations to solve pair-wise and multi-view fine registrations.
  - The construction of a novel deterministic outlier detection function, independent of the relative pose or distance histograms between two partial surfaces, to deal with partial overlap between patches during registration.
  - The use of variational implicit functions for surface and texture integration. VIS (surface function), VTF (texture function), VWF (weight function).

  - The use of a common surface representation throughout the whole algorithm such that integration and registration can be merged into a single step.

- Application of the proposed surface registration framework to establish dense point correspondences on inter-subject facial surfaces. To robustly solve the facial surface registration, we propose to apply the presented robust statistical surface registration framework based on a non-rigid thin-plate-spline (TPS) transformation model and a similarity measure deducted from a VIS target surface representation. Outliers, due to partial overlap or missing data, are dealt with using a deterministic outlier detection function and probabilistic outlier-process, similar to the outlier handling during fine registrations of patches. Because of the non-rigid and complex nature of the registration problem a deterministic annealing optimization is applied.

- The development of a complete statistical approach for computer-based craniofacial reconstruction. The core most abstract formulation of the presented approach is: *What is the most plausible or probable face, according to a statistical craniofacial model being a face distribution, given the (erroneous) skull data.*

  - The construction of an elaborated database for craniofacial surface purposes using non-invasive measuring technology.

- Statistical modeling:

  - The construction of a combined statistical craniofacial model. In order to eliminate the template-related model bias and to minimize the unrealistic character of the reconstructions caused by large generic model deformations, we propose a flexible statistical craniofacial model for reconstruction. We model the combined population-dependent variance and covariance of complete skin surface shape, 52 anatomical face landmarks with tissue depths, property (age, BMI and gender) values and skin surface gray-value texture information.
  - Statistical manipulation according to skull property values. The creation of more realistic sub-population craniofacial models according to a set of skull properties, derived from a anthropological examination, is obtained by modeling and removing facial variations originating from property differences in the facial database. The result is a property normalized craniofacial model.
  - The generation of more lifelike reconstructions based on a texture map derived from facial geometry, property value and tissue depth information. This is accomplished trough the incorporation of gray-valued texture variation.

  - The application of the proposed surface registration framework to establish model-based point correspondences on facial surface samples. These model-based point correspondences are compared with the TPS-based point correspondences to validate correspondence consistency and therefore the quality of the statistical model.

- Statistical reconstruction:

  - The application of the surface registration framework to tackle the skull registration to obtain robustness against noise and outliers.
  - The construction and the embedding of two different skull representations into the registration framework. The first is novel and based on VIS representation and the second is a cranio-metric representation augmented with surface normal information.
  - The derivation and the embedding of four different craniofacial models into the registration framework. Eliminating subsequently covariance between different information parts from the proposed model leads to more traditional craniofacial models working with a single facial template in combination with generic deformations.
  - The initial establishment of a craniofacial reconstruction validation framework based on a database with known facial outlooks using quantitatively and qualitatively facial comparison techniques.

Part I

Application and general methodology

Chapter 2

Craniofacial modeling and reconstruction

"Sometimes it is believed that the old ways become obsolete, but the future is nearly a repetition and mimicking of the past"

2.1 Manual CFR

When confronted with a corpse that is unrecognizable (figure 2.1) due to its state of decomposition, skeletonisation, mutilation or incineration, and if no other identification evidence is available, craniofacial reconstruction can be considered. The goal of craniofacial reconstruction (CFR) is to recreate an estimate of the face of an individual at the time of death, starting from the skull specimen. Hopefully, this will trigger recognition by relatives such that further identification evidence can be gathered from a restricted list of candidates. Although craniofacial reconstruction is a valuable tool in the initiation of the process of identification, eventually, positive identification has to be obtained by classical techniques such as radiographic and dental comparisons or DNA-analysis.

All facial reconstruction techniques are based on the assumed relationship between the soft tissue envelope and the underlying skull substrate [8]. Several 3D manual methods for facial reconstruction have been developed and are currently used in practice. These reconstructions consist of physically modeling a face on a skull replica (the target skull) with clay or plasticine. The Russian anthropologist Gerasimov (1971) [9] was the first to make a manual reconstruction by modeling the complete anatomy of muscles and soft-tissues covered by a thin layer of skin onto the skull. This anatomical or "Russian"technique, also referred to as the *morphoscopic* technique, was further refined by Lebedinskaya and co-workers [10]. In the mean time an alternative technique was developed in the United States, called the *morphometric* method [11]. This technique consists of building the soft tissue layers in bulk, without much re-

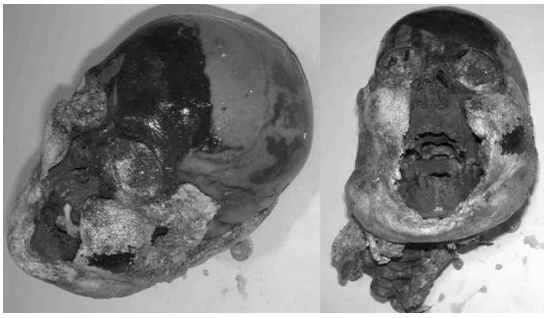


Figure 2.1: Example skull to identify. Due to the severe state of decomposition craniofacial reconstruction is to be applied to help the investigation out of an impasse.

gard to the rest of the underlying anatomy, approximating tabulated average tissue depths at a sparse set of landmarks on the skull and interpolating in between. An example of this technique is depicted in figure 2.2. More recently, Richard Neave [12] used both Russian and American methods laying the foundation for the *combined* technique, which was further developed by Caroline Wilkinson [13] and her team at Manchester. The facial interpolation in between manually placed soft tissue depth markers is based on the facial muscles and soft tissue depths used in an anatomical manner. An example of this sophisticated reconstruction technique is shown in figure 2.3 on the same skull of figure 2.2. Proponents of this method believe that since the faces are reconstructed according to the rules of anatomy, artistic subjectivity in areas with limited tissue depth measurements is reduced.

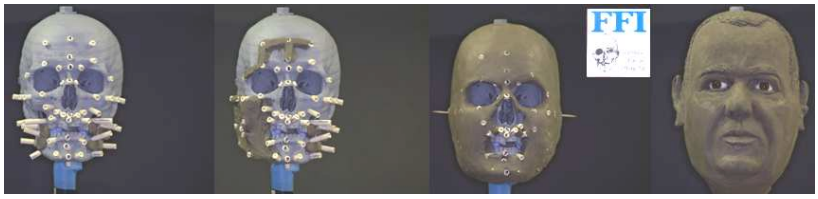


Figure 2.2: Example manual craniofacial reconstruction, using the morphometric technique performed by the FFI department of the Belgium federal Police (Ben Claes)

Manual reconstruction methods, however, require a lot of anatomical and artistic modeling expertise and remain as a result subjective. Two different artists have a different feeling or knowledge about the facial features to reconstruct and make two different faces, which can be seen by comparing the final faces in figures 2.3 and 2.2, using in this case a slightly different manual technique. According to Davy *et al.* [14] this point is further illustrated in [15], in which multiple facial reconstructions of several victims from the Green

Figure 2.3: Example manual craniofacial reconstruction, using the combined morphoscopic and morphometric technique performed by Ludo Vermeulen (Belgium)

River serial-killer were created. The results were highly variable between practitioners and with little success. Furthermore, these reconstructions take a lot of time (several days), and, hence, are often limited to a single reconstruction.

Computer-based methods, on the other hand, can be consistent (given the same input data and modeling assumptions, the same output results) and objective (knowing all the modeling assumptions). Moreover, since these methods can be executed in a short time, multiple reconstructions from the same skull using different modeling assumptions (older, thicker, ...) can be obtained. A final advantage of using virtual computer models is the enhanced possibility of visualization. The skull and the reconstructed face can be visualized together by making the face transparent, such that the face-skull relationship can be examined and, if necessary, corrected. The development of software for computerized facial reconstructions of an individual would be of benefit to various law enforcement agencies, by allowing faster, easier and more efficient generation of multiple representations of an individual.

In the following section we will give an exhaustive and critical review of the published methods for computerized craniofacial reconstruction. We structure the discussion following the general work-flow of a computer-based reconstruction set-up. Specific choices for each of the components in this work-flow are compared, leading to the proposal of a new, improved computerized reconstruction method that will constitute the central part of this thesis.

2.2 Computerized CFR: taxonomy and critical review

Current computerized techniques all share the same general work-flow shown in figure 2.4. They are essentially a virtual mimicking of manual reconstruction techniques and can be compartmentalized into six components. The skull specimen is first examined by anthropological experts to determine subject properties such as age and gender. These properties parameterize the craniofacial reconstruction model which codes for the a priori knowledge about facial shape and its link to the skull substrate. In order to execute a computerized re-

construction, a virtual skull copy has to be digitalized from the real specimen. During reconstruction the craniofacial model is fit to a compatible computer representation of the virtual skull. In a final stage, the reconstructed facial shape can be additionally textured and rendered in order to generate images for further distribution.

2.2.1 Anthropological examination

The skull, which has a craniometric individuality as distinctive as a fingerprint [16], is the essential substrate and source of information of any craniofacial reconstruction. In a first instance, an anthropological examination is performed in order to determine a set of skull properties including age, gender, race and stature [17]. Sometimes it is even possible to determine or estimate the Body Mass Index (BMI) based on remaining soft tissue layers on the skull or additional evidence found at the crime scene (clothes e.g). Computer assistance for estimating race and gender is given by the software package FORDISC 2.0 [18, 19] which has recently been upgraded to FORDISC 3.0 [20]. This interactive computer program offers custom discriminant functions for up to 21 manual, human-guided, cranial measurements. Unlike most previously published functions, this program allows classification, even with incomplete remains for which only a limited number of measurements is possible.

Although no completely automatic computer-aided method supporting such examinations exists too date, it is likely that automated skull classification procedures will be developed in the same way facial archetypes describing a various cohort of people are being developed today [21].

2.2.2 Skull digitalization

In a second part of the work-flow, a digitized version of the skull is obtained in order to translate the skull shape into a machine-readable format for further processing and visualization. This is comparable with a cast or mold construction of the skull for manual reconstruction techniques in order not to destroy the original skull and, therefore, possible evidence.

Pioneering work on computerized three-dimensional craniofacial reconstruction was performed by Vanezis *et al.* (1989) [22]. In their work and in [14, 23, 24, 25] the skull is digitized using a laser scanning system. This system consists of a computer controlled rotating platform, a projected laser line and a video camera interfaced to the computer. Simple triangulation is used to derive the 3D shape of the skull from the recorded digital video images.
Thanks to recent advances in medical imaging technology, Computer Tomography (CT) scanners have become a useful alternative for acquiring a digital copy of the skull. A CT scanner produces a set of stacked 2D slices through the skull on which the hard-tissue structure can be clearly distinguished from other structures based on their relatively high CT intensity values. The surface of the 3D skull can be extracted from the 2D slices using marching cubes [26], hysteresis thresholding [27] or dynamic contour models [28]. All recent

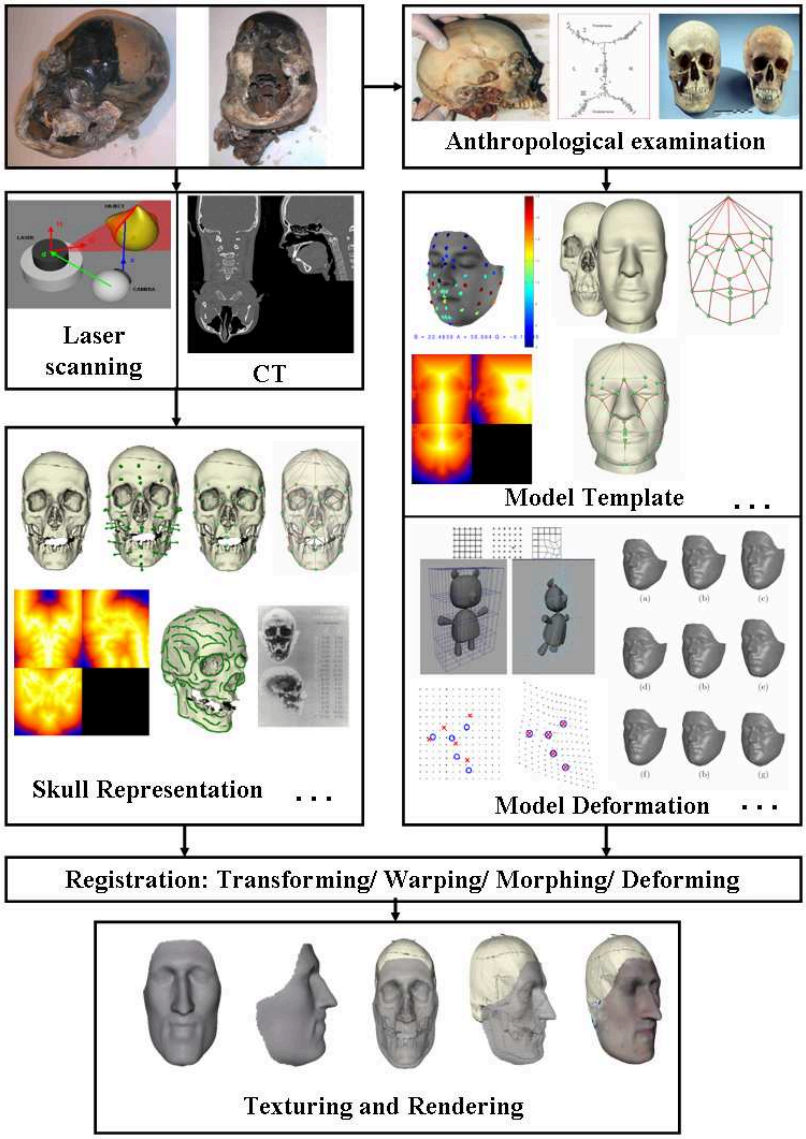


Figure 2.4: General work-flow of computerized craniofacial reconstruction techniques

computerized reconstruction techniques use CT scanners for skull digitalization.

Both laser-based and CT-scanners have their limitations and shortcomings, especially in the context of skull surface acquisition. CT scanners are sensitive to amalgam teeth fillings resulting in heavy streak artifacts in the images as shown in figure 2.5. The 3D surface extraction based on the stacked 2D CT images generates a kind of moustache artefact around the teeth. Although laser scanning is a general technique for scanning the outer-surface of 3D objects, the skull surface may be too complex to reconstruct based on laser-line projections. More importantly both scanners have a limited scanning resolution, such that very small details of the skull are not acquired or copied. One has to keep in mind that, while working with a virtual copy of the skull, errors due to the used acquisition technique are introduced and are to be dealt with. A virtual copy is never an exact copy of the skull specimen.

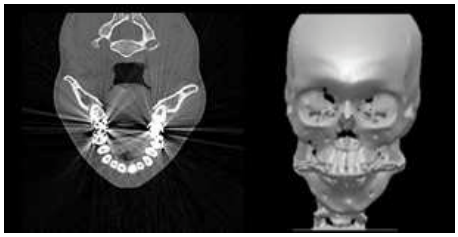


Figure 2.5: Streak artifacts in CT images resulting from amalgam teeth fillings

2.2.3 Craniofacial model

An essential step in the computerized reconstruction work-flow is the definition of a craniofacial model,which codes for the knowledge about human facial shapes and their relationship to the underlying skull as learned over a representative database of living subjects. A craniofacial model contains three components: (1) a facial template describing the facial outlook, (2) a facial template information source containing the knowledge relating faces to skulls and, (3) a geometric deformation model describing the class of transformations allowed in fitting the craniofacial model to a given skull specimen.

Craniofacial model template

A craniofacial *model template* is a computer representation of the shape of a human face and can be classified following the hierarchical taxonomy shown in figure 2.6. At the top level, templates either represent a holistic, complete view of the face or represent the face as a collection of parts like the nose, mouth, ears e.g. [22, 29, 30] and even muscles [14]. Complete facial templates describe either a single reference head or multiple reference heads. A single reference head can either be a generic/average face [23, 31, 32, 33, 34, 35] or

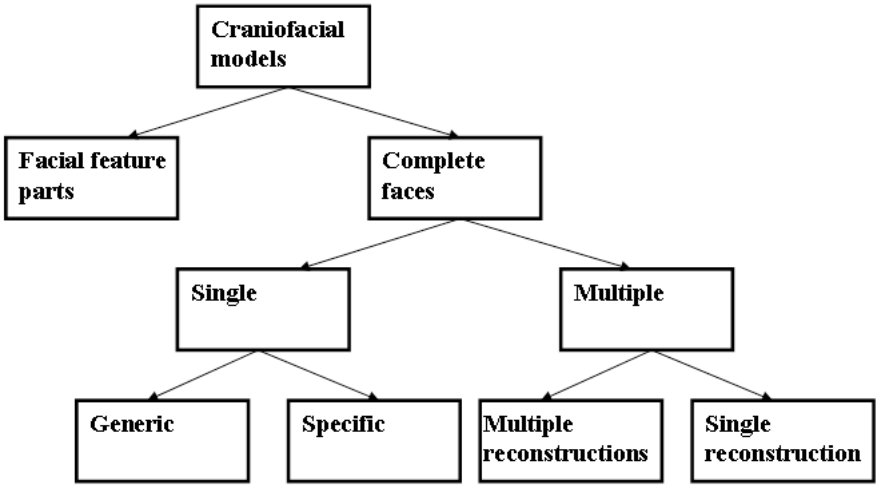


Figure 2.6: Classification of craniofacial model templates

a specific similar face in terms of age, gender,race and skull dimensions drawn from a database [24, 25, 28, 36, 37, 38, 39, 40, 41, 42, 43].

A first way to cope with multiple templates is to make an individual reconstruction starting from every template in the database resulting in multiple reconstructions for one skull substrate. Afterwards these results can be combined as a weighted average [44, 45, 46] or analyzed using a technique from structural analysis, Principal Component Analysis (PCA) [47]. A second, dual, way is to analyze the templates with PCA [3, 4, 5, 6, 7, 48, 49, 50] before the deformation towards the skull and to make a single reconstruction based on the analyzed knowledge of multiple faces.

Property Normalization. By selecting appropriate template(s),and the way multiple model templates are integrated, the reconstruction can be more or less tailored to the skull property estimates as obtained from the anthropological examination. We call this tailoring process *property normalization*.

- In a *pre-reconstruction property normalization mode*, the craniofacial model is normalized to the skull properties before a reconstruction is made. This is the preferred mode, since the use of a craniofacial model with properties or attributes that differ too much from the target skull, could make the reconstruction too difficult, if not impossible.

- In a *post-reconstruction property normalization mode*, only the reconstructed results are normalized to the estimated properties. This mode of operation is interesting when a property of the reconstruction needs to be altered from its value at the time of death to generate, e.g., a younger

version of the victim when he/she could have been missing for several years prior to his/her death.

In [25] a simple linear geometrical interpolation between two reconstructions with different skull properties, is used to generate intermediate reconstruction results. They also state that ageing is a non-linear process, because ageing will progress at different rates for different people affecting different parts in the face in different ways.

A typical way to tailor the craniofacial model is to select one or a set of reference heads from a database, of which the properties are similar to the ones of the skull [24, 25, 28, 36, 37, 38, 39, 40, 41, 42, 43]. However this requires an extended database containing enough samples for every possible sub-population, which is labor intensive to acquire. Alternatively, one can try to model or learn geometrical facial variations originating from attribute differences between faces in a database in order to simulate age, corpulence, race and gender changes. The advantage is that under-sampled sub-populations in a database can still be represented by the interpolating nature of the attribute modeling or learning process. In [4, 6, 7] a linear statistical interpolation of attribute differences is learned. Claes *et al.* [4] use this to generate multiple reconstructions starting from one reconstruction in a post-reconstruction mode, while Claes *et al.* [6, 7] use this to manipulate the craniofacial model itself in a pre-reconstruction mode. Before the fitting or reconstruction process, shape variation in the statistical model coded by the values of the given properties is removed. This results into a property normalized database of facial entries with the same property values as the skull specimen, from which a new statistical model with property variation eliminated can be calculated and used during reconstruction. In other words every face in the database is projected into the sub-population according to the given skull properties by simulating an appropriate amount of age, gender and BMI change per face. A non-linear ageing model is constructed in [51] based on the same statistical face representation as in [4, 6, 7]. However, learning a non-linear ageing model requires more and better sampled facial entries in a database compared to a linear ageing model, which can already give a satisfying simulation.

Model Bias. If inappropriate templates are chosen, *model bias* can occur. Model bias is the leaking through of specific facial feature details, originating from the model template, into the reconstruction. Using only a single generic or specific best look-alike reference head, the potential to produce model-biased reconstructions is high, which is illustrated in figure 2.7. Indeed, when using a subject-specific best look-alike template, based on similarity in ancestry, gender and age, unwanted facial features of the template remain visible in the final reconstruction (red ellipses in figure 2.7). Using a generic face template, on the other hand, results in too smooth and unspecific a reconstruction. To reduce or to resolve the model bias it is better to work with multiple reference heads. Thinking about the human artist being the craniofacial model for manual reconstruction techniques it is logical to work with multiple templates.

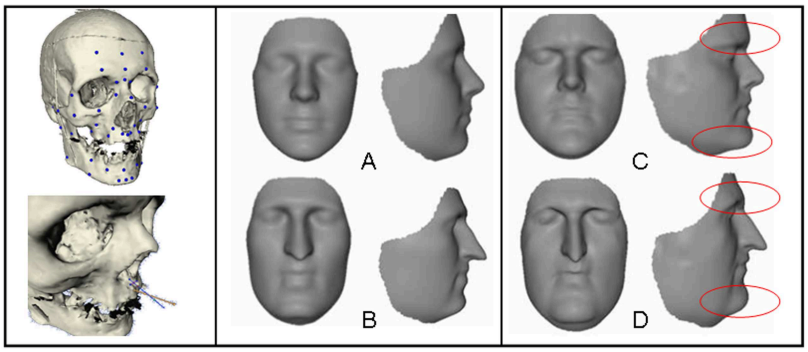


Figure 2.7: Illustration of model bias. Two reconstructions (B,D) made of the same skull, represented as 52 skull landmarks and an estimate of the nose tip, using a single generic (A) and specific (C) model template in combination with a TPS deformation.

The knowledge of the artist is based on seeing many faces during his or her live. An artist knowing only the geometry of his/her own face will make reconstructions resembling him/herself. The goal is to create a computerized craniofacial model keeping objectivity and incorporating the same amount of knowledge as a human artist, which can only be done based on a database of many faces.

Craniofacial model template information

The facial *template information* describes the knowledge relating faces to skulls and, independent of the model template class, incorporates facial surfaces, tissue thicknesses, skull surfaces and/or facial muscles. The simpler ones incorporate only facial surface information [22, 23, 29, 30, 31, 32, 33, 41]. Capturing skull surface information combined with facial surfaces is typically obtained using CT scans of living or deceased subjects [28, 35, 36, 37, 38, 39, 40, 42, 44, 45, 46, 49, 50, 52, 53]. A generic face model consisting of the outer facial surface and 24 facial muscles was build by [34], using graphics modeling techniques. This gives the possibility to animate the final reconstruction based on muscle movements. Mang *et al.* [54] use Magnetic Resonance Imaging (MRI) scanners to build up a database of reference heads. The major advantage of MRI compared to CT is the enhanced visualization of different soft-tissue types (muscles, fat,... e.g.). However, the distinction between hard-tissues and air is less clear in MRI, making it hard to extract skull information from those images. Finally, in our work [3, 4, 5, 6, 7, 48] a combined facial surface template is used with soft-tissue thicknesses measured at 52 anatomical landmarks according to [1, 2].

The choice of *model template information* is dependent on the type of scanning material used to scan living subjects for building the craniofacial model. The major advantage of using a CT scanner to build up a database is the pos-

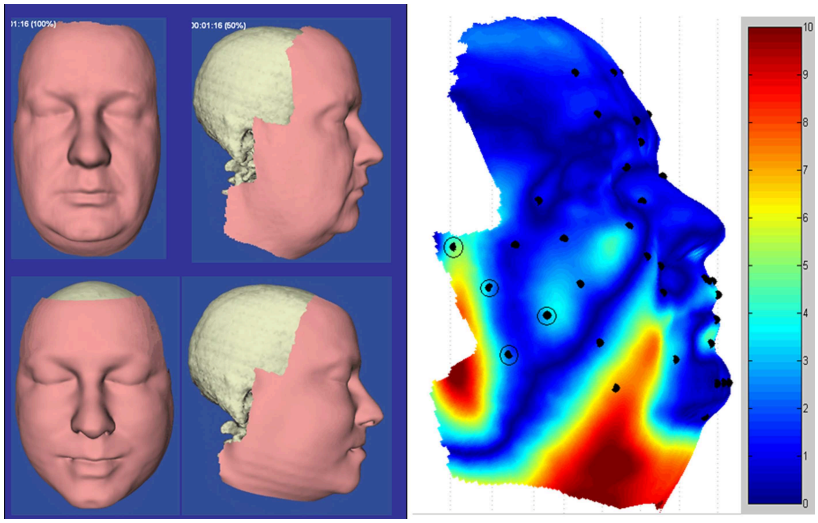


Figure 2.8: The facial geometry of a person in a horizontal supine position (bottom row) and an upright position (top row). Local surface differences visualized by means of a color-code ranging from 0 to 10 mm (right image)

sibility to have dense skull representations in combination with dense related tissue depth measurements. A disadvantage however, is the involved irradiation dose during CT scanning, limiting the reference database to patient data and incomplete head scans or deceased subject data. Furthermore, CT images are acquired of subjects in a horizontal, supine position. As a result, due to gravitational forces, facial shapes extracted from CT images will differ from the typical facial shape as viewed in standard upright position, which is illustrated in figure 2.8. Recently, Cone-Beam CT scanners have been developed and are being used in practice in which the subject can be scanned in an upright position. However the signal to noise ratio in such scanners is much lower than traditional CT scanners resulting in noisy 2D CT slices, making it harder to extract the skull surface accurately. An alternative to CT scanners is the use of MRI scanners, which are considered not to be harmful. Again dense soft-tissue depth measurements can be acquired, including differentiation between different types of soft-tissue (muscles, fat e.g.). However, the same remark concerning the difference in facial outlook between supine and upright position of the subject during scanning stands for MRI scanners. Furthermore, the link between the soft-tissue measurements of a subject and the underlying skull is hard to establish because of the poor hard-tissue visualization in MRI. A third choice of scanning material is the use of laser based or photogrammetry based scanners. These scanners give the possibility to scan a person in a an upright position, without being harmful. Laser scanners are considered to be accurate, but the required acquisition time is in the order of several seconds making it difficult for a person to stand still during acquisition, influencing the quality

of the scan. Photogrammetry scanners are less accurate but have a fast acquisition time, which becomes important when scanning children. Both types of scanners are limited to scan the outer facial surface, such that additional measuring technology (ultrasound e.g.) is required to measure tissue-depths. Because of the manual labor involved measuring the tissue-depths, the number of measurements is limited to a sparse instead of a dense set of anatomical landmarks on the face.

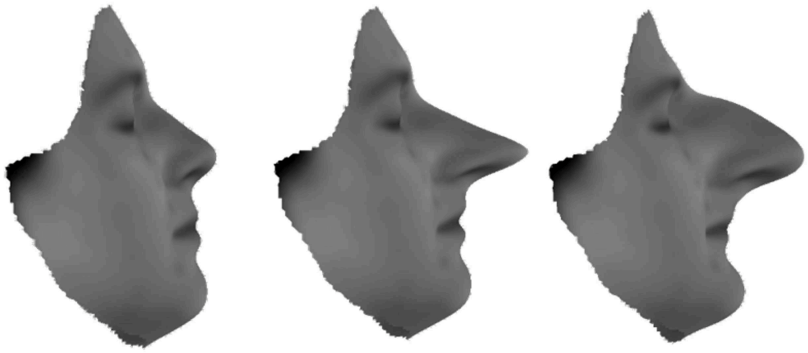


Figure 2.9: Nose Tip manipulation of a face (left) with a generic TPS based deformation (middle) and a face-specific statistical PCA based deformation (right)

Craniofacial transformation model

The *transformation models* which are applied to the facial templates in order to geometrically align them with the given skull are very diverse.
A first class of rigid transformations apply only a translation and rotation. Incorporating scale and skew as well leads to another class of affine transformations. Rigid or affine transformations are used by all techniques in order to bring both the skull and the craniofacial model into the same coordinate system.
Starting from this rough alignment or crude registration, local surface deformations are needed in order to adapt the model template exactly to the skull specimen by making use of non-rigid or non-affine deformation models.

Generic deformations The majority of computer-based methods make use of generic non-rigid deformations, which can be applied to a broad range of objects besides faces. A popular non-rigid and smooth transformation mechanism is the use of Radial Basis Functions (RBF) as in [41]. RBF's are functions centered around points in which the deformation or displacement is known, such that the deformation is nicely interpolated and extrapolated in between and outside the known displaced points or features. Thanks to the interpolation and extrapolation properties other points besides the known displaced points

of the facial template can be deformed or displaced as well. Thin plate Splines (TPS) [55] are used as radial basis functions in [28, 35, 38, 39, 43, 47, 54], while [14, 31, 32, 33, 46] use B-splines defined on a regular grid. [53] uses 2D instead of 3D bicubic splines, because they converted the template deformation problem from 3D to 2D by using a cylindrical mapping of the skull. [42] also applies a 2D warp or deformation after taking a 2D image of the frontal view of the 3D reference skull. Other generic non-rigid deformation models used in computerized reconstruction techniques besides RBF's are non-uniform scalings [22], digital cosine transformations [44, 45], polygon-based deformations [29, 56], diffused scattered motion fields [40], trilinear transformations [57], volume distortion functions based on disc fields [36] and volume deformation based on local cylindrical coordinate systems [37].

The advantage of using generic transformations or deformations is that they are applicable to a wide range of objects and that no training or deformation learning phase is necessary. However care is to be taken when using generic deformations, because no knowledge of facial geometry and/or anatomy is incorporated. They are just "smooth", but can deform a face into a Picasso or Pinocchio look-alike face when not used carefully. To illustrate this we applied an exaggerated displacement to the tip of the nose of the face in figure 2.9 (left). Subsequently the face was deformed based on this displacement making use of a generic TPS interpolating deformation (figure 2.9 (middle)). The resulting face has a bigger, but very pointy Pinocchio alike nose, which is unrealistic and implausible according to human anatomy. No problem arises when the differences between the model template and target skull are small. However, when these differences are relatively large, the required deformation will be more pronounced, which can result in unrealistic, caricature-like or implausible facial reconstructions. This is also visible in figure 2.7, where large differences between several corresponding landmarks on the model templates and the target skull (the nose tip e.g) generate caricature-like final outlooks of the reconstructions due to the large amount of non-affine TPS deformation that has been applied. Especially, when using the subject-specific face, the nose of the reconstruction looks unrealistic.

Face-specific deformations The use of a face-specific instead of a generic non-rigid deformation model was first proposed by Claes *et al.* (us) in [3] and also used by [49, 50]. Here, the facial template and the deformation model are interrelated into a statistical model consisting of a geometrical averaged facial template and a correlation-ranked set of modes of principal variations that capture the major changes between different facial outlooks in a database centered around the average. This template can be considered as an elastic mask of which the elasticity is defined as the statistically allowed correlated variation or covariance of facial surfaces. By changing the statistical model parameters between the statistically determined boundaries, the mask is deformed in a face-specific way only. The disadvantage of using a statistical model is the need of an elaborate database incorporating enough variability. In order to overcome a lack of elasticity of the mask due to a limited database

or limited variation in the database, [4, 5, 6, 7] combine the face-specific deformations with a more generic TPS based deformation, resulting in a flexible but constrained deformation of the mask.

Generating a face-specific deformation model typically includes a learning phase or a statistical analysis of a facial database. The advantage compared to generic deformations is that faces can be deformed in a face-specific manner within statistical boundaries, guaranteeing the facial plausibility. To show the difference we deformed the face in figure 2.9 (left) based on the same nose tip displacement making use of a statistical interpolating deformation extracted from a PCA facial model resulting in the face of figure 2.9 (right). The final nose is anatomically more acceptable than the one based on the TPS transformation. The disadvantage is that the modes of face-specific deformation learned from a database are entirely dependent on the samples in the database. Having a small database or a database with low inter-subject variance, generates a small and too restrictive set of deformation modes, so that faces a-typical to the database are hard to reconstruct. It is best to have an extended database or otherwise to combine the face-specific deformations with generic deformations in order to reconstruct a wide range of different faces.

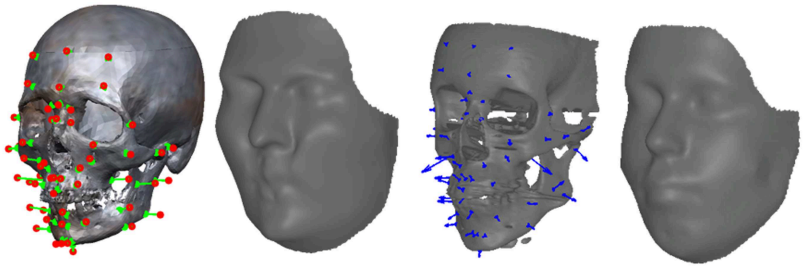


Figure 2.10: Two examples of errors in the skull representation (52 virtual dowels, based on manually indicated landmarks and surface normal information) generating surface inconsistencies of the final reconstruction. Both reconstructions are made based on an adjusted average facial template (in terms of skull properties) in combination with a generic TPS based deformation

2.2.4 Unknown skull representation

The fourth component in the work-flow is the skull representation, which is related to the choice of craniofacial model. A skull can be represented or parameterized in different ways depending on the type of relationship between soft- and hard-tissues or template information incorporated in the craniofacial model. The skull representation for manual reconstruction techniques is a physical copy of the skull surface combined with dowels in specific anatomical or craniometric landmarks whose lengths are equal to average tissue depths according to the anthropological examination results.

Craniometric skull representation

Quite a few computer-based methods [14, 22, 23, 29, 30, 31, 32, 33, 34, 40, 41, 43, 54, 56] adapt the manual reconstruction skull representation by interactively positioning virtual dowels in a sparse set of anatomical landmarks on a virtual copy of the skull whose lengths equal tabulated average tissue depths. Others [14, 40, 58] expand the number of virtual dowels in between the manually placed dowels at mathematically calculated intermediate points. This virtual dowel representation is typically used in combination with a model template consisting of facial surfaces, without skull surface or tissue depth information except for [43, 54]. The end-points of the virtual dowels represent estimates of landmarks on the face surface, which are typically indicated on the facial model template as well. Alternatively, some methods [22, 30, 33, 56] use a tissue growth algorithm starting from the virtual dowel representation to create an initial featureless mask of the face lacking eyes, ears, nose and mouth. The tissue thickness distributions known in the landmarks are grown starting from the skull surface by interpolating the thickness values for points on the skull surface. This kind of fundamental facial shape representation can be used in combination with facial models consisting of separate facial feature parts. Something similar is done by [14], but instead of growing tissue, they model the muscles onto the skull after which the facial parts and skin are added. [3, 4, 5, 6, 7] also uses a sparse set of anatomical landmarks indicated manually on the skull, but no thickness values are set out. Instead soft-tissue thicknesses are incorporated into the craniofacial model instead of the skull representation in order to exploit the influence of tissue thickness on facial geometry and vice versa. Thicknesses are set out starting from landmarks on the skin surface perpendicular to the skin surface, resulting in 52 skull landmarks per face in a database. We refer to these skull representations based on or starting from manually indicated craniometric or anatomical skull landmarks as sparse *craniometric skull representations*.

Dense point skull representation

Having a model template with both facial and skull surface information, a *dense point skull representation* is possible. A first possibility is to represent the skull with a set of control or feature points [28, 37, 42, 50] which can be determined automatically and which can have a denser distribution than the sparse craniometric representations discussed in the previous paragraph. In the most extreme scenario, every point on a discrete digital representation of the skull surface is used, but in practice a subset of points is defined. A second possibility is the use of points on crest-lines as in [38, 39, 47]. These lines follow the salient lines of the skull surface like the mandible, the orbits, the cheekbones or the temples [39]. Thirdly, [35, 53] convert the 3D skull into a 2.5D range image representation using a cylindrical mapping, which is a conversion of (x, y, z) cartesian coordinates into angular θ, height h and radial r coordinates. A 2D image is then defined such that the horizontal axis

corresponds to the angular coordinate θ, the vertical axis corresponds to the height coordinate h and the intensity in every pixel of the image is set to the radial coordinate r. After the 3D to 2.5D conversion, a sparse set of salient or fiducial points are manually indicated onto the 2.5D images.

Implicit Skull Representation

A completely different skull representation is used in [44, 45, 46]. A signed distance transform (sDT) of the skull is constructed, representing for each element in a 3D matrix the shortest euclidian distance to the skull surface, zero on the surface, positive inside and negative outside. This *implicit skull representation* is a dense almost continuous representation and does not only code the original skull surface, but, at the same time codes an infinite set of surfaces inside and outside the skull at certain iso-distances from the skull, which are smoothed versions of the original surface.

2.2.5 Model to skull registration

The fifth component describes the registration, also called fitting or matching, which is the process of finding and applying the geometrical relationship or transformation between the craniofacial model and the skull. Starting from a skull representation that is compatible with the chosen cranio-facial model, a similarity criterium is defined, which expresses the goodness of fit of the craniofacial model to the skull. The craniofacial transformation model is combined with the similarity criterium into an objective function. During registration a set of transformation model parameters is searched for, optimizing the objective function, by increasing the similarity criterium or goodness of model fit within a range of possible transformation parameters. Once these parameters are known, the complete facial template can finally be deformed/warped/morphed towards the skull specimen.

A frequently used similarity criterium is the distance between corresponding points or features in between the skull representation and the craniofacial model. These correspondences can either be known or unknown before registration. If the correspondences are known, the registration simply reduces to finding the parameters of the chosen deformation model that transform the known corresponding points onto each other leading to an optimization of the similarity criterium within the objective function. In case correspondences are unknown they need to be searched for as well during registration leading to a more complicated optimization task.
Having a *sparse* set of craniometric landmarks indicated both on the skull and the model template in 3D [3, 4, 5, 6, 7, 23, 29, 31, 32, 34, 40, 41, 43, 54] or 2.5D [35, 53], correspondences are established manually and are therefore known before registration. The registration simply reduces to transforming the manually determined corresponding landmarks onto each other. Using the set of determined parameters the complete facial template can be transformed towards the skull. However, having a more *dens* skull representation, the prob-

lem of finding corresponding points or features is hard to solve manually and some automatic correspondence procedure is needed. [28] and [40] both start from interactively manual indicated points in order to establish more corresponding points automatically. However, no further details are given. [37] uses a volume-based correspondence algorithm, but again no further details are given. [42] uses a correlation based measure in order to find corresponding points in between the target skull and a reference skull. A very popular automatic registration procedure with unknown correspondences is the Iterative Closest Point (ICP) algorithm [59], which converges monotonically to the nearest local minimum of the objective function expressing remaining distances between corresponding points. ICP is a iterative registration algorithm alternating between correspondence update and parameter update in every iteration step, until convergence occurs. A corresponding point on a target surface of a point on a reference surface is the point lying closest in terms of distance. Based on the closest points being corresponding points the transformation parameters are updated, until the change in objective function falls below a certain accuracy threshold. [38, 39] use an ICP alike algorithm to align the points of the skull crest-lines. [47] also uses ICP to align crest-lines on skulls, but afterwards more corresponding points besides the points on the crest-lines are searched for, based on distance and point normal information. [57] extends the ICP objective function with a matching symmetry constraint in order to find more correct dense correspondences between skulls. In [50] they further developed their method using these dense correspondences to find the parameters of a statistical face-specific deformation model as a missing data problem.

When using an *implicit* skull representation, no explicit point correspondences are searched for, instead the complete sDT, represented as a 3D image, of a reference skull is matched or registered with the sDT of the target skull based on a sum of squared differences (SSD) similarity criterium in [45] or a Mutual Information (MI) similarity criterium in [46]. The result is a better control of the deformation away from the original skull surfaces, because skull surface information is extended in terms of iso-distance surfaces to the skull.

An important issue concerning the *registration* of the craniofacial model towards the skull is the presence of small errors or noise and gross errors or outliers within the skull representation on which the registration is performed. Besides the fact that a virtual copy is never an exact but a noisy copy of the skull specimen, additional errors are introduced during skull representation build up. Making a registration robust against small errors is done by incorporating a global transformation regularization into the objective function, restricting the deformation. Instead of exactly interpolating or obeying the skull representation an approximation is allowed, resulting in smoother (TPS) or more facial plausible (PCA) deformations. The amount of regularization needed is typically dependent on the amount of noise present in the skull representation. The type of regularization is dependent on the a-priori knowledge incorporated in the craniofacial transformation model. Outliers, however are gross errors compared to the noise level present in the majority of the data

constituting the skull representation and they severely influence the objective function and therefore the final transformation parameters. Increasing the regularization to deal with outliers makes the deformation too restrictive such that the model template is not allowed to change. Instead outliers are to be detected and either removed or their influence on the deformation parameter estimation is to be reduced. This can be accomplished by using individual skull data confidence values incorporated into the similarity criterium part of the objective function. The lower the confidence value the lower the influence on the parameter estimation. Removing outliers is then similar to setting the appropriate confidence values equal to zero.

Manually indicating anatomical or craniometric landmarks on a skull is error prone even for the most experienced anthropologist. Furthermore, the surface normal in a point of the skull surface needed to set out an according tissue-depth is only an approximation and heavily dependent on the skull surface sampling density. Ignoring these errors during registration can result in surface inconsistencies of the final reconstruction. Two examples are depicted in figure 2.10. The mouth region of the left reconstruction is incorrect due to bad placement of the virtual dowels in that region. Small unrealistic peaks and dips are visible in the cheek and chin region of the reconstruction on the right. A first challenge is to determine the amount of regularization needed. A second challenge is to deal with wrongfully indicated landmarks. Using an interactive program to correct the virtual dowels manually makes it possible to generate a reconstruction based on trial and error. However, this can be very laborious and time consuming. Alternatively, an individual confidence value per landmark can be assigned and the deformation parameters can be determined accordingly as in [54] for TPS based deformations. In practice, it is not easy to manually assign confidence values, again leading to a trial and error approach. So far, no automatic estimation of the noise level determining the needed amount of regularization and of the individual confidence parameters is incorporated during registration for craniofacial reconstruction.

Not only manually indicated skull representations incorporate errors. Similar errors in points or features extracted with an automatic procedure to construct and register skull representations exist. Here also, regularization is to be incorporated during registration and outliers are to be dealt with. Additionally, a new kind of outlier points or features can occur, which are skull points or features extracted from the target skull having no corresponding feature or point in the craniofacial model or vice versa. This kind of errors due to missing correspondences are to be detected and removed during registration. [39, 47] incorporate such outlier removal during registration of the target skull crest-lines and model crest-lines.

2.2.6 Texturing and Rendering

Generating a good approximation of the geometry of the face belonging to a skull substrate is the most important prerequisite of a craniofacial reconstruction technique, but in order to generate a lifelike appearance *texturing*

the reconstruction is necessary. Manual reconstructions can be refined by the artist painting directly onto the clay model. A few current computerized reconstruction techniques apply texturing in a final step of the work-flow after the geometry of the reconstruction is created. 3D modeling software can be used to virtually paint the eyes and mouth e.g. onto the 3D surface of the reconstruction. Another possibility is to use texture mapping as in [14, 39], which is a process akin to applying wallpaper to a flat surface. In this case the wallpaper is a facial texture map like in figure 2.11 (b), mapped onto the 3D surface of a face (figure: 2.11 (a)), resulting in a lifelike 3D face viewable from different angles (figure: 2.11 (c)). Besides texturing, a 2D sketch of the 3D reconstruction can be generated to make the final result more alive as in [39].

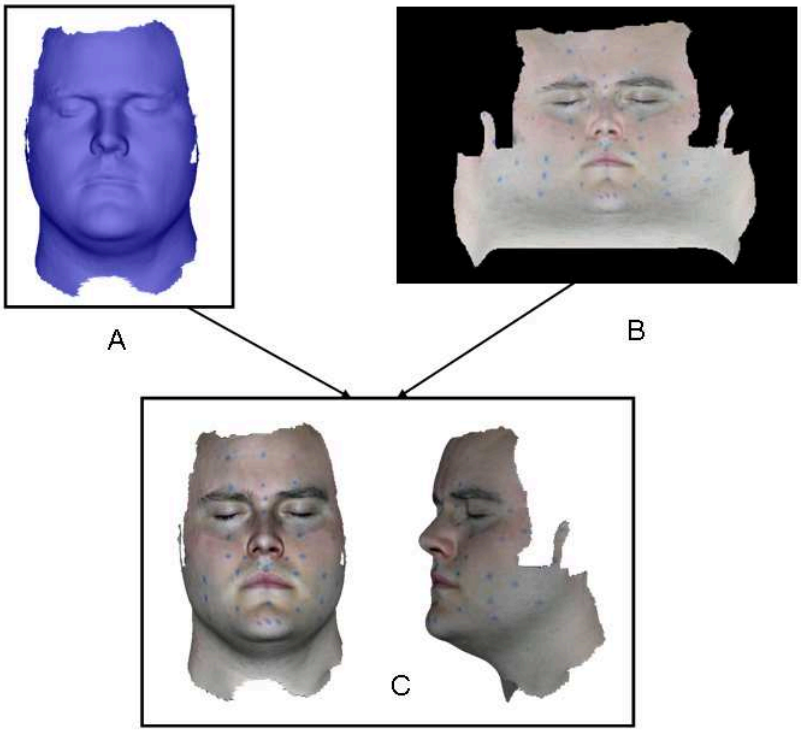


Figure 2.11:

The final textured reconstruction result can now be *rendered* or visualized with proper 3D visualization software. The advantage is that both the skull and the facial outcome can be rendered together, while playing with a transparency parameter in order to examine the face-skull relationship. This kind of visualization makes it interesting to evaluate the reconstruction and to correct possible errors.

Texturing the reconstruction is not without warning, because it can trigger

a recognition into the wrong direction. Applying the texture map of a certain living subject onto the facial geometry of a different person will trigger the recognition towards the former person. It is not advised to use a texture map from a particular individual, introducing unwanted fine detail. Instead an average texture map should be created and used to render the 3D reconstruction. Average texture maps, belonging to different sub-populations are to be created in the same way a craniofacial model for different sub-populations is necessary. So far texturing is considered a separate final step after constructing the facial geometry of the reconstruction. However, it is thinkable of incorporating texture information besides skull, face and tissue thickness information into a craniofacial model based on multiple faces. This could give the possibility to generate more specific compared to averaged texture maps according to the geometry of the face, because the relationship between texture and geometry could be modeled in the same way the relationship between skull properties and geometry is modeled in [6, 7]. The use of 3D modeling software capable of painting onto a 3D surface is interesting to create particular details like scars or birthmarks when they are known to be present.

2.2.7 Validation

The last issue in the design of a computerized reconstruction technique is *validation*. From an engineering point of view, a developed method can work fine, but is useless in practice until properly validated. A first attempt to set up a validation framework for craniofacial reconstruction methods has been defined in [6]. A proper validation framework will hopefully substantiate on a scientific basis the added value of reconstruction methods during crime-scene investigations.

A database of skulls with known facial outlooks is needed for a proper validation. Validation can be based on a *leave-one-out cross-validation* scenario. Here every skull in the database is removed in turn, and used as a test case. The resulting facial skin surface of the reconstruction technique can then be compared with the skin surface of the test case representing the ground truth. In a first instance a quantitative error evaluation can be performed by observing local surface differences. Calculating the distance between every point on the reconstructed skin surface and its closest point on the ground truth skin surface is a typical way to do this. However, this kind of distance calculations are an underestimation of the local errors [60, 61], because they are typically lower than distances calculated between anatomically corresponding points onto the two surfaces. Comparing the two surfaces this way is interesting for evaluating the reconstruction performance in terms of accuracy and provides a spatial map of the difficulty of each facial region to be reconstructed. However, the final goal of craniofacial reconstruction is not reconstruction accuracy, but rather recognition or identification success. [6] used Euclidian Distance Matrix descriptors [62] for 3D surfaces to compare the reconstruction with a database of faces including the ground truth. Several other computer-based face recognition descriptors can be used, which is currently an active research topic [63].

A more realistic, human subjective, identification process can be simulated by generating face-pool tests. Given an image of a reconstruction and a set of possible candidate images extracted from the database, a human observer is asked to indicate the face from the face-pool most similar to the given reconstruction. However, it is very difficult to generate realistic face-pool tests. Finally, comparing the performance of different craniofacial models, skull representations and/or registration algorithms is to be done within the same test set and with the same comparing techniques in order to be objective.

2.3 Conclusion: towards an improved computerized craniofacial reconstruction method

Forensic facial reconstruction aims at estimating the facial outlook associated to an unknown skull specimen for victim identification. Manual reconstruction methods require a lot of anatomical and artistic modeling expertise and are as a result subjective and time consuming. The development of software for computerized facial reconstructions of an individual would be of benefit to various law enforcement agencies, by allowing faster, easier and more efficient generation of multiple representations of an individual. Computer-based methods are essentially a virtual mimicking of manual reconstruction techniques involving an anthropological examination influencing the choice of craniofacial model. Subsequently this model is deformed or transformed towards a virtual copy of the skull based on a similarity criterium extracted from a skull representation. The proper transformation parameters are determined by optimizing an objective function combining the similarity criterium and the craniofacial transformation model.

Besides the issues about craniofacial reconstruction in general, a number of choices are to be made and a number of implementation issues are to be dealt with while designing a computerized reconstruction technique. For every step in the general work-flow of figure 2.4 several choices exist and problems are to be coped with.

Based on the critical review in section 2.2, we propose in this thesis a complete *statistical* approach for computer-aided craniofacial modeling and reconstruction. The philosophy behind the reconstruction technique is to find the most plausible or probable face, according to a face distribution or model, given the erroneous skull data. A statistical craniofacial model for reconstruction is used, eliminating the model-bias and minimizing the unrealistic character of the reconstructions caused by large generic model deformations, incorporating the combined population-dependent variance and covariance of complete skin surface shape, 52 anatomical face landmarks with ultrasound tissue depths, property (age, BMI and gender) values and skin surface gray-value texture information, calculated from an extended database of faces measured in an upright position. Furthermore, statistical manipulation or simulation of properties is used to fine-tune the model towards a given set of skull properties.

In order to build up the statistical craniofacial model and to use it for cran-

iofacial reconstruction, several *registration* problems are to be tackled. Firstly, the complete facial surface acquisition of database samples is based on assembling partial surface acquisitions. Secondly, different facial surface entries within the database need to share dense correspondences in order to perform a statistical shape analysis. Thirdly, the resulting craniofacial model is to be transformed towards a skull substrate. Despite the three different registration goals, the developed and used registration strategy is very similar. In the next chapter we explain the core registration methodology or strategy using a statistical or probabilistic robust objective function formulation combined with implicit surface representations.

Chapter 3

Surface registration methodology

3.1 Introduction

In this chapter we will present a general framework for robust surface registration. Surface registration is defined as the process of finding the geometrical transformation between two or more surfaces that aligns the surfaces as well as possible. Independent of the surface representations, these transformations map the coordinates of each point of one surface, called the *floating* surface, into the coordinate space of the *target* surface.

As mentioned at the end of the previous chapter, different surface registration tasks are to be solved as part of the statistical craniofacial modeling and reconstruction procedure proposed in this thesis.

- First, the acquisition of complete facial surfaces that make up the database requires the assembling of partial surfaces. Indeed, most three-dimensional (3D) surface acquisition systems generate several partial surfaces that are to be combined to obtain a complete 3D object surface. The reason is that a number of these acquisition systems can only compute 3D information within a limited field of view. A 3D acquisition from a single viewpoint is referred to as a patch or partial surface. A complete surface is built from many patches or partial surfaces and can be seen as a patchwork. The required number of views is mainly determined by the complexity of the object. A more complex object will require more views at which point it becomes interesting to have a fully automatic complete surface acquisition system based on partial acquisitions.
  Combining several patches into a single surface, involves two main phases depicted in figure 1.2. First, in the registration phase, the patches are accurately aligned into a common coordinate frame. Secondly, in the integration phase, the registered patches are integrated into a single entity. During patch registration, the floating surface is a particular partial

surface or patch, while the target surface can either be another particular patch or an integration of all the other partial surfaces having a partial overlap with the floating surface. The *patch registration* must be robust against noise, both statistical measurement errors as well as gross reconstruction errors, present in the individual patches. Furthermore, it must be able to cope with outliers defined as points on one patch that have no corresponding point on the other patch due to partial overlap between the different patches.

- A second registration task concerns the establishment of dense correspondences between different complete facial surfaces. The determination of dense point correspondences is required to further statistically analyze the facial shapes. Point correspondences are obtained by mapping 3D points with known connectivity defined on a carefully constructed generic reference face onto every face in the database as shown in figure 1.3. Here, the generic reference face is the floating surface while the faces in the database are the target surfaces. The result is that every facial surface in the database is represented by the same amount of points with the same connectivity, such that for every point on one surface the corresponding point on every other surface in the database is known.
  The *facial surface registration* must be robust against noise and must be able to cope with outliers due to partial overlap or missing data. Noise is here defined as incomplete or even erroneous intermediate point correspondences as estimated during the iterative process of registration. Indeed, as we will see later on, the registration is calculated in an alternating sequence of determining point correspondences and calculating the geometrical transformation mapping the estimated corresponding points onto each other. Especially at the beginning of the registration procedure, erroneous point correspondences will be established since the shapes of the floating and the target surface are still very different. Missing data occur because the 3D surface acquisition device is not able to acquire 3D information from highly textured regions (such as facial hair), resulting in 3D surface acquisitions with holes or missing data. An example is visible in figure 1.3: due to the presence of side-whiskers, the 3D facial surface contains holes or misses 3D surface information at the location of the whiskers.

- The third and most important registration task is the craniofacial model to skull registration illustrated in figure 1.4. Here, the floating surface is the craniofacial model template, while the target surface is the unknown skull surface. The *skull registration* must be robust against noise or small errors and must be able to cope with outliers or gross errors. Smaller errors are due to skull digitization effects and errors in manually indicating craniometric landmarks (if necessary). Gross errors occur in the teeth region because of CT-reconstruction artefacts. Furthermore, gross errors are also caused by inconsistencies between the manual indication of the cranial landmarks (during Ultrasound-based measuring of

soft tissue thicknesses on the subjects in the reference database) and the automatic determination of these landmarks during the computerized reconstruction procedure.

In order to solve these registration tasks, we developed a **unifying framework for robust, statistical, surface registration**. It uses a principled, statistical approach to simultaneously handle the measurement noise and outliers. The framework is presented in extenso in section 3.3. Prior to this, in section 3.2, we propose a continuous, smooth and analytical, **implicit surface representation**, which we will show to have particularly interesting properties for registration purposes. This framework is extensively used throughout the thesis in the context of computerized craniofacial reconstruction. Starting from the theoretical foundations of the surface registration framework given in this chapter we will tackle every registration problem that is encountered. Specific details about the different registration problems are given in chapter 4, 5 and 7 for the patch, complete facial surface and skull registration problems respectively. However, the framework is also suitable to solve other surface registration problems, besides the ones we encountered, as well.

3.2 Implicit surface representation

A surface can be mathematically defined as a two-dimensional manifold in three-dimensional Euclidean space $\mathbb{E}^3$. The most familiar surface examples are those that arise as the boundaries of solid objects. Surfaces can be either closed or open. An open surface is a 2D manifold consisting of a surface interior, which is always non-empty, and its complement the surface boundary. Open surfaces are not closed in either direction. This means that moving in any direction along the surface will cause an observer to hit the edge or boundary of the surface. The patch and complete facial surfaces in figures 1.2 and 1.3 are open surface examples. A closed surface is one that is boundaryless, having an empty boundary, and compact. Moving in any direction on such surfaces will cause the observer to travel forever without hitting an edge. The skull surface in figure 1.4 and the spherical surface in figure 3.1 are closed surface examples.

For the purpose of registration a proper surface representation is to be chosen. Three major surface representation classes exist: mesh or point-based surfaces, parametric surfaces and implicit surfaces.

- In virtual reality and computer graphics or animation, a surface is often represented by a *surface mesh* of node points or vertices connected by edges, forming simple polygons such as triangles or quadrilaterals. The process of creating a mesh is called tessellation and generates a discrete surface representation. Connected vertices or meshes are very suitable for visualization purposes and are used throughout figures 1.2,

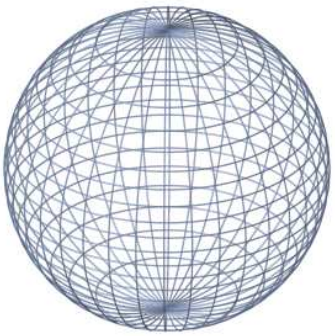


Figure 3.1: A wire-frame visualization of a mesh represented sphere. A parametric and implicit representation of the same sphere are given in (3.1) and (3.2) respectively.

1.3 and 1.4. Similarly, the sphere visualization in figure 3.1 is also obtained by rendering the vertices and the polygon edges of the spherical mesh representation. Often this representation is derived from another representation as an intermediate or final representation for surface visualization purposes. The disadvantage of a discrete representation is that the surface is approximated by discrete samples forming local polygons and therefore normals and tangents to the surface are approximated as well and cannot be determined analytically.

- Instead, it is better to use a continuous surface representation. One such continuous surface representation is the parametric representation. Let f be a continuous, injective function from $\mathbb{R}^2$ to $\mathbb{R}^3$. Then the image of f is said to be a *parametric surface*. The parametric representation for the sphere in figure 3.1 is:

$$\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned} \tag{3.1}$$

  where θ and ϕ are the 2D angular parameters of the representation. Tessellation for visualization is obtained by altering the parameter values within predefined boundaries generating node points or vertices on the surface. Restricting the range of the angular parameters for the sphere in figure 3.1, generates an open surface instead of a closed surface, where the boundary is defined by the parameter range. The disadvantage of parametric surfaces is their topology dependency.

- Implicit surfaces, on the contrary, are topology independent continuous representation. Suppose that f is a smooth function from $\mathbb{R}^3$ to $\mathbb{R}$ whose

gradient is nowhere zero[1]. Then the locus of zeros of f is defined as the *implicit surface*. The implicit representation for the sphere in figure 3.1 is:

$$f(x, y, z) = x^2 + y^2 + z^2 - r^2 = 0 \tag{3.2}$$

Implicit functions define the surface of an object as the zero level set ($f(x, y, z) = 0$) of an analytical function $f(x, y, z)$. The 3D surface is therefore embedded into a 4D euclidian space $\mathbb{E}^4$ consisting of 3D point positions and an additional functional value. The advantage of representing surfaces implicitly is similar to the "flat-landers"problem (2D creatures not being able to see similarities in mirrored versions of each other in a 2D world, but easily in a 3D extension). Indeed, by adding an additional dimension, some representational disadvantages of parametric representations disappear. Most notably, the representation automatically adapts to the topology of the objects surface. Note that the function f in (3.2) is a signed Distance Transform (sDT)[2] to the spherical surface with radius r, which is a special kind of implicit function. For points on the spherical surface the function value is zero. When a 3D point $\mathbf{r} = (x, y, z)$ lies inside/outside the spherical surface the function value $f(\mathbf{r})$ is the negative/positive distance to the spherical surface in the direction of the sDT gradient $\nabla f(\mathbf{r})$ in that point. This is a very useful and interesting property for surface registration purposes. When the points of the floating surface are evaluated into the sDT of the target surface, a notion of remaining distance and updating direction is implicitly given within these points. 3D target surface information is extended to 4D information in terms of iso-distance surfaces, such that the space between the floating and the target surface is filled with useful information about directed distances to the target surface. Tessellation of implicit sDT's for visualization purposes is typically obtained using marching cubes [26] or marching triangles [64].

Because of its continuous, topology-independent nature and the interesting properties for registration purposes we propose to use a sDT (or an approximation thereof) implicit representation of the target surface. In the remainder of this thesis, we refer to an implicit surface as being the sDT. The surfaces to start from are, most often, mesh-represented surfaces, acquired with a 3D capturing device. Therefore, an implicit function is to be derived from this mesh represented surface by fitting this function through the points on the surface.
The implicit function in (3.2) is an algebraic implicit function and is easily determined because of the relative simple spherical geometrical form of the surface in figure 3.1. However, *algebraic implicit functions* represent a limited class of surfaces depending on the geometrical complexity of the objects.
A first way to deal with this problem is to sample the functional on a regular

[1] Otherwise the surface, defined as the locus of zeros of f, cannot be extracted as a two-dimensional manifold.

[2] To be more exact: a signed squared distance transform.

volumetric grid instead of determining the functional analytically. A volumetric grid is placed around the surface and subsequently the signed distance of every grid point to the closest point on the surface is determined and stored. However, these *grid-based implicit surface* representations are very memory inefficient. A surface is by definition very thin, such that the grid resolution in the vicinity of the surface must be high, while the resolution further away from the surface can be much lower. A grid structure is not intrinsic to a surface, but artificial. Furthermore, sDT values and gradients are only approximated by discrete grid point value interpolations.
Alternatively and more interesting than the grid-based sampled implicit functional, analytical *variational implicit functionals* of complex surfaces can be derived using scattered data interpolation or fitting techniques, where a smooth function is created that passes through a given set of data points. The smoothness is imposed by variational techniques.

We applied this variational fitting framework to create implicit surface representations for the target surfaces. Details about the variational implicit functions are given in section 3.2.1. The construction of variational implicit surfaces based on these variational implicit functions is explained in 3.2.2. Furthermore, the same variational implicit function fitting technique is used in section 3.2.3 to determine an interior-exterior or boundary function for open surfaces to cope with the lack of boundary notion within implicit surface representations.

3.2.1 Variational implicit functions

A variational implicit function (VIF) is created from constraint points using a variational scattered data interpolation approach. These functions are obtained as special cases of a scattered data interpolation problem using variational techniques, which we briefly introduce here [65]. Given a collection of constraint points $\{\mathbf{r}_i = (x_i, y_i, z_i) |\ 1 \leq i \leq N\}$ with function constraints $\{h_i |\ 1 \leq i \leq N\}$, we want to construct a "smooth"scalar-valued function $\mathfrak{f}(\mathbf{r})$ such that $\{\mathfrak{f}(\mathbf{r}_i) = h_i |\ 1 \leq i \leq N\}$. The scattered data interpolation can be formulated as a variational problem where the desired solution is a function, $\mathfrak{f}(\mathbf{r})$, that minimizes a functional $\|\mathfrak{L}(\mathfrak{f})\|^2$, representing integrated curvature e.g, subject to the interpolation constraints $\mathfrak{f}(\mathbf{r}_i) = h_i$. The solution of the variational implicit interpolation function $\mathfrak{f}(\mathbf{r})$ can be expressed as a sum of weighted radially symmetric basisfunctions (RBF) centered at the constraint locations.

$$\mathfrak{f}(\mathbf{r}) = v(\mathbf{r}) + \sum_{i=1}^{N} \lambda_i \phi(\|\mathbf{r} - \mathbf{r}_i\|) \tag{3.3}$$

with:

$$\mathfrak{f}(\mathbf{r}_i) = h_i \qquad i = 1, \ldots, N \tag{3.4}$$

where:

- $v(\mathbf{r})$ is a low degree polynomial, typically linear or quadratic.
- ϕ is a real valued, radially-symmetric, radial basis function (RBF) centered around a constraint point.
- the λ_i's are the RBF coefficients or weights.
- the $\mathbf{r}_i$'s are the RBF centers or constraint point locations.

Let $V = \{v_1 \dots v_l\}$ be a monomial basis for polynomials of the degree of $v(\mathbf{r})$, and $\mathbf{c} = \{c_1 \dots c_l\}$ be the coefficients that give $v(\mathbf{r})$ in terms of this basis. The constraints (3.4) can then be written in matrix form as a linear system (3.5) for determining Λ and $\mathbf{c}$, and hence $\mathfrak{f}(\mathbf{r})$.

$$\begin{pmatrix} \Phi & V \\ V^t & 0 \end{pmatrix} \begin{pmatrix} \Lambda \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} h \\ 0 \end{pmatrix} \tag{3.5}$$

where:

$$\begin{array}{ll} \Phi_{i,j} = \phi(\|\mathbf{r}_j - \mathbf{r}_i\|) & i,j = 1,\dots,N \\ V_{i,j} = v_j(\mathbf{r}_i) & i = 1,\dots,N \quad j = 1,\dots,l \end{array} \tag{3.6}$$

Depending on the type of regularization functional $\|\mathfrak{L}(\mathfrak{f})\|^2$, solving (3.5), which is called RBF fitting, for more than a few thousand constraint points is computationally expensive and impractical (due to the size of the RBF kernel Φ). Algorithmic improvements have been proposed involving hierarchical domain decomposition [66] and fast multi-pole methods [67] for the function fitting and evaluation to reduce the original computational complexity from $O(N^3)$ to $O(N \log N)$ and the storage requirements from $O(\frac{N^2}{2})$ to $O(N)$. This makes RBF's practical for 1 million and even more constraint points. The approximation algorithms are implemented in the FastRBF^{TM} library, which is commercially available at [68]. As a consequence of the approximation methods, the equalities of the constraints (3.4) are only approximated. Therefore two user-defined accuracies are needed. The first is the fitting accuracy and is defined as:

$$\max_{i=1,\dots,N} |\mathfrak{f}(\mathbf{r}_i) - h_i|) \tag{3.7}$$

The second is the evaluation accuracy defined as:

$$\max_{i=1,\dots,N} |\mathfrak{f}(\mathbf{r}_i) - a_i|) \tag{3.8}$$

Where a_i are the approximate values of the RBF's at the points $\mathbf{r}_i$ for $i = 1,\dots,N$.

For the RBFs ϕ, we use 3D biharmonic splines which are known as smoothest interpolators, in the sense that they minimize certain energy functionals and interpolate the data. Given a set of constraint points $\{\mathbf{r}_i = (x_i, y_i, z_i) | \, 1 \le i \le N\}$ with function constraints $\{h_i | \, 1 \le i \le N\}$, the function $\widehat{\mathfrak{f}}(\mathbf{r})$ that satisfies the interpolation conditions and minimizes a bending energy regularization functional $\|\mathfrak{L}(\mathfrak{f})\|^2$:

$$\widehat{\mathfrak{f}} = \operatorname{argmin}(\|\mathfrak{L}(\mathfrak{f})\|^2) \tag{3.9}$$

where

$$\begin{aligned}\|\mathfrak{L}(\mathfrak{f})\|^2 = \int_{R^3} (\tfrac{\partial^2 \mathfrak{f}}{\partial x^2})^2 + (\tfrac{\partial^2 \mathfrak{f}}{\partial y^2})^2 + (\tfrac{\partial^2 \mathfrak{f}}{\partial z^2})^2 + \dots \\ 2(\tfrac{\partial^2 \mathfrak{f}}{\partial x \partial y})^2 + 2(\tfrac{\partial^2 \mathfrak{f}}{\partial x \partial z})^2 + 2(\tfrac{\partial^2 \mathfrak{f}}{\partial y \partial z})^2 d\mathbf{r}\end{aligned} \tag{3.10}$$

is the 3D biharmonic or thin-plate spline (TPS) $\phi(\mathbf{r}) = \mathbf{r}$. The RBFs will smoothly change in value between the constraint or interpolation points. As a consequence interpolation and extrapolation are inherent to the $\mathfrak{f}$ functionals.

Using TPSs, a smooth transition of the function $\mathfrak{f}$ in between consecutive constraint points is obtained, while interpolating through the function constraints. When dealing with noisy constraint points and function constraints, additional smoothing is necessary and, instead of exact interpolation, an approximation of the constraint points is obtained. In [69] different possible smoothing techniques for VIFs during evaluation or fitting are explained and shown on noisy data. One of those techniques, called spline-smoothing, is achieved during fitting by modifying the smoothness criterion in (3.9):

$$\widehat{\mathfrak{f}} = \operatorname{argmin} \left(\nu \|\mathfrak{L}(\mathfrak{f})\|^2 + \frac{1}{N} \sum_{i=1}^{N} (\mathfrak{f}(\mathbf{r}_i) - h_i) \right) \tag{3.11}$$

with $\nu \geq 0$ and $\|\mathfrak{L}(\mathfrak{f})\|^2$ the smoothness penalty defined in (3.10). According to [69], ν is similar to an inverse spring constant. If $\nu = 0$, the springs are inextensible and the VIF $\mathfrak{f}(\mathbf{r})$ is forced to pass through the data points. When $\nu > 0$, the springs can stretch and the VIF is pulled close to the function constraints in the constraint points, but not forced to pass through them. Solving the fitting problem in (3.5) with $\nu > 0$, again leads to a closed form linear system:

$$\begin{pmatrix} \Phi + \nu I & V \\ V^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} \Lambda \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} h \\ 0 \end{pmatrix} \tag{3.12}$$

with I the identity matrix. Increasing the value of ν is essentially trading off constraint fidelity or confidence in favor of smoothness or approximation instead of interpolation.

3.2.2 Variational implicit surfaces

A variational implicit surface (VIS) is an iso-surface of a 3D variational implicit function (VIF), as was first introduced in [70] for closed surfaces. Alternatively, a VIS representation is the representation of a surface with a variational implicit function. Given the machinery of variational techniques for scattered data interpolation, how can we use this to build a VIS representation? We follow the recipe as proposed by Carr *et al.* [65] and illustrated in figure 3.2. The constraints in (3.4) that we use to build a VIS $= \mathfrak{d}(\mathbf{r})$ representation, consist of surface location-constraints and normal-constraints. The location-constraints are the 3D coordinates of the surface points through which the zero iso-surface or level-set of the function $\mathfrak{d}(\mathbf{r_i})$ should pass ($h_i = 0$) and the

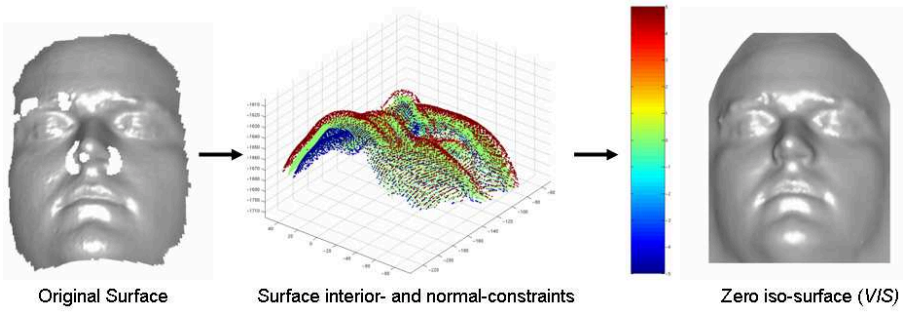


Figure 3.2: Construction of a VIS representation. Starting from the original surface (left) interior- (green points) and normal- constraints)blue and red off-surface points) are defined (middle), through which a variational implicit function is fitted resulting in the VIS representation being the zero level-set of the VIF (right). The latter is visualized after tessellation of the VIS

normal constraints are positive/negative valued constraints ($h_i = \pm d$) that are placed near the location-constraints, and are positioned outside/inside the surface at a certain distance d along the normal direction. Choosing these constraints leads to an implicit function $\mathfrak{d}(\mathbf{r})$ that is a very good approximation (at least close to the surface, and gracefully degrading away from it) of a Euclidean signed distance-to-surface function or sDT.

When using three-dimensional biharmonic or thin plate splines as RBFs, interpolation and extrapolation are inherent to the VIS representation. Small holes in the surfaces are interpolated and the extrapolation properties allow the function $\mathfrak{d}(\mathbf{r})$ to be evaluated anywhere in 3D, on and off the surface. Evaluating the function in a point off the surface will give an approximation of the signed distance to the surface in that point when using the location- and the normal-constraints. In figure 3.2 the interpolation and extrapolation properties for an example surface are clearly visible.

The fitted 3D variational implicit function, from which the zero iso-surface $\mathfrak{d}(\mathbf{x}) = 0$ can be extracted using implicit function tessellation techniques, is a continuous function that can be evaluated anywhere, not just on a discrete grid. Furthermore, gradients are continuous and smooth and can be determined analytically. The representation, like any other implicit representation, is topology independent, but in contrast to most other implicit function representations, no memory inefficient and resolution limiting volumetric-grid is used. The number of parameters to be stored is only dependent on the number of constraint points and can even be reduced using approximation techniques. These approximation techniques also allow for a built-in data smoothing with variational implicit functions and surfaces either during evaluation or during fitting of the function (3.3), like the spline-smoothing in (3.11) during fitting. In [69] the different smoothing techniques during evaluation or fitting are explained and shown on noisy data.

input triangles	X dim.	Y dim.	Z dim.
54850	211 mm	298 mm	260 mm
GRID	grid sampling	Dimensions	memory
representation	1 mm	221x249x182	**78.25** Mb
VIS	accuracy	constraints	memory
representation	1e-04 mm	55790	**1.85** Mb

Table 3.1: Comparison of memory occupation between an uniform grid and a VIS representation.

Besides smoothness properties and the continuous analytical formulation, an additional advantage of VIS representations compared to grid-based implicit surface representations is memory efficiency. Table 3.1 shows a comparison of memory occupation in matlab^{TM} R13 (double accuracy) between a uniform grid representation with no memory optimization and a VIS representation of a facial surface. Observing table 3.1, note that, in practice the structure of grid representations are optimized (e.g Octree structures) instead of uniform, reducing the grid memory occupation but leading to additional overhead.

3.2.3 Variational boundary functions

VIS representations were originally proposed in [70] for closed surfaces. However, due to the implicit nature of the VIS representation, being embedded in a higher-dimensional space, it is hard to retain the boundary concept for open surfaces. As a result, open surfaces are extrapolated outside their boundaries during tessellation of the corresponding implicit function depicted in figure 3.2 (right). These extrapolations and interpolations are correct in the vicinity of the boundary, but deteriorate as one moves further away from the boundary. In order to use VIS representations correctly for open surfaces such as the patch and complete face surfaces in figures 1.2 and 1.3, we developed an additional variational implicit function. This function implicitly encodes the boundary of an open surface, such that the interior and boundary or exterior concept for open surfaces is retained within a VIS representation. This variational boundary function (VBF) is constructed using the same variational interpolation or fitting technique from section 3.2.1 based on distance to border or boundary constraints.

The determination of the boundary constraints is illustrated in figure 3.3. First, we determine the border ∂S of an open surface S. Then, the geodesic distance[3] $d(\mathbf{r}_i, \partial S)$ of every point $\mathbf{r}_i$ belonging to the interior $\exists S$ of the open surface S to ∂S is determined. We then define a set of boundary constraint points $\{\mathbf{r}_i | i = 1, \ldots, N^s\}$ with N^s the number of points on the surface S and

[3]The geodesic distance between two points is approximated as the shortest list of connected edges (of the tesselated surface) connecting the two points.

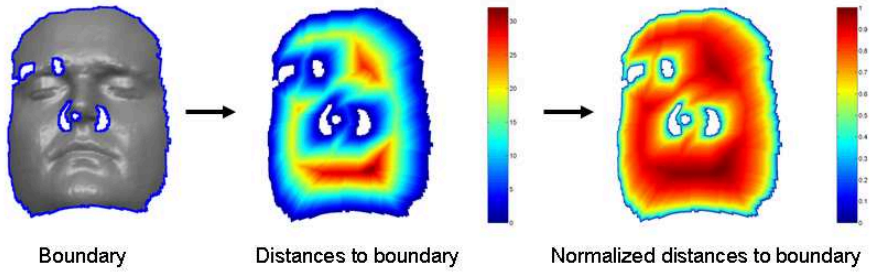


Figure 3.3: Construction of the boundary constraints. First, the boundary points of the surface are detected (blue points in left figure). Then the Euclidean distance of every point to the closest boundary point is determined (middle). Finally, these distances are normalized between 0 and 1 according to (3.13) (right). Distances are visualized using a colormap

associated function constraints $\{h_i | i = 1, \ldots, N^s\}$:

$$h_i = w' \log(1 + w'' d(\mathbf{r}_i, \partial S)) \tag{3.13}$$

This function has been chosen such that the function constraints increase rather quickly for points further away from the border of the patch. Furthermore, distances to the boundary are normalized between zero and one, but any other similarly behaving function can be substituted. Possible values for (w', w'') are $(1/\log(100), 99/(\max_i d(\mathbf{r}_i, \partial S)))$.

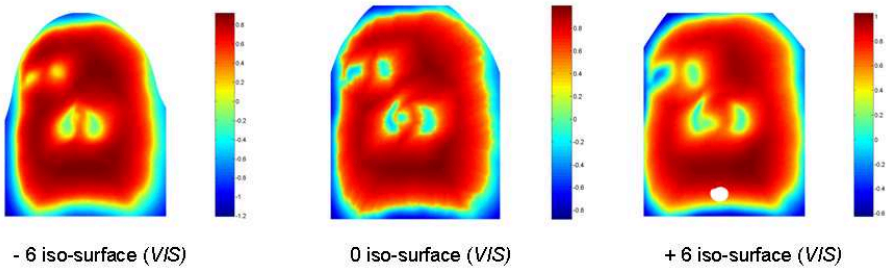


Figure 3.4: Evaluation of the VBF in three different iso-surfaces of the VIS representation of the surface in figure 3.3 (Left) tessellation of the - 6 mm iso-distance surface evaluated in the VBF, (middle) zero iso-distance surface tessellation and (right) + 6 mm iso-distance surface tessellation

If we construct a variational implicit boundary function VBF $=\mathfrak{b}(\mathbf{r})$ from the constraints in equation (3.13), we obtain a partitioning of $\mathbb{E}^3$ into a positive and negative region. Based on the interpolation and extrapolation properties of $\mathfrak{b}(\mathbf{r})$, points in 3D located directly above and below the surface S will evaluate positive in $\mathfrak{b}(\mathbf{r})$, decreasing slowly with increasing distance from the surface,

while points next to and far away from the surface will evaluate negative. Indeed, starting in a point in the middle of S and approaching the border of S, the function $\mathfrak{b}(\mathbf{r})$ will gradually decrease to zero. When traversing the border this decrease will continue and $\mathfrak{b}(\mathbf{r})$ becomes negative. This can also be seen in figure 3.4. Observing the middle image in figure 3.4, if we evaluate the tessellation or mesh of the zero iso-distance surface from the VIS representation in the VBF and only retain the tesselated vertices evaluating zero or positive (red to green values), while discarding the tesselated vertices evaluating negative (light blue to dark blue), then we reduce or cut-off the extrapolations generating a surface tessellation close to the original open surface. Through the adaption of the cutting off threshold from zero to a small negative number, an amount of extrapolation and interpolation can be retained if wanted. Using the concepts of VISs and VBFs we build a surface registration framework in the following section and subsections.

3.3 Robust surface registration framework

In this section we present our general methodology for point sampled surface registration, based on a statistical embedding which is robust to noise and outliers and which uses implicit surface representations for the target surface as presented in the previous section. Robustness to noise is obtained by casting the registration as a statistical estimation problem resulting in a Maximum A-Posteriori (MAP) objective function formulation. As stated in [71], MAP approaches have proven to be very useful theoretical frameworks, equipped with a set of powerful and well developed mathematical techniques with clear assumptions. Robustness to outliers is obtained by defining a weighted MAP objective function using inlier and outlier beliefs resulting from appropriate probabilistic inlier- and outlier-processes. Throughout this section, the floating surface, in contrast to the target surface, is represented as a set of connected 3D points. The final goal of registration is to reposition the points of the floating surface as good as possible at prescribed distances (most often zero, but some applications require an offset distance) to the target surface.

As already partially mentioned in the introductory chapter, a registration algorithm generally requires the following four main components.

- A *similarity measure* (section 3.3.2), which estimates the correctness of a given transformation by comparing the transformed floating surface with the target surface.
- A *transformation model* (section 3.3.1) restricting the geometric transformation to an application-specific family of transformations.
- An appropriate error, energy or *objective function* (section 3.3.3), which combines both the similarity measure and the transformation model.
- A *numerical optimizer* (section 3.3.4) which generates a sequence of transformation estimates converging to the position where the similarity

between both surfaces is maximal within the range of possible transformations by optimizing the objective function towards a global optimum.

Often, prior to surface registration, an initialization (section 3.3.5) is required to bring both the floating and target surface within the same coordinate system and to ensure that the optimizer converges to the correct global optimum of the objective function.

3.3.1 Transformation model

The transformation model stipulates how the points on the floating surface can be transformed to the target surface. Given a floating surface $S^f = \{\mathbf{r}_i^f | i = 1, \ldots, N^f\}$ consisting of N^f three-dimensional points. Suppose the transformation can be abstractly represented by a function $\mathfrak{T}$ with parameters $\boldsymbol{\theta}$. Applying it to a floating surface point $\mathbf{r}_i^f$ will map or transform it to its new location $\mathbf{r}_i^{f*} = \mathfrak{T}(\mathbf{r}_i^f, \boldsymbol{\theta})$ such that the complete transformed floating surface S^{f*} can be defined as:

$$S^{f*} = \mathfrak{T}(S^f, \boldsymbol{\theta}) = \{\mathfrak{T}(\mathbf{r}_i^f, \boldsymbol{\theta}) | i = 1, \ldots, N^f\} \tag{3.14}$$

The choice of using a particular type of transformation model reflects our prior knowledge on the registration problem. Depending on the registration problem and the presence of noise within the target surface, not all transformations are feasible or realistic so that certain mapping constraints are imposed to ensure that the transformation behaves according to our prior knowledge. These constraints, inducing a transformation regularization, can be modeled probabilistically using a Gibbs prior distribution on the parameters $\boldsymbol{\theta}$, restricting the space of possible transformations:

$$p(\boldsymbol{\theta}) = \frac{1}{Z_{part}} \exp\left(-\nu \|\mathfrak{L}(\boldsymbol{\theta})\|^2\right) \tag{3.15}$$

This Gibbs prior distribution expresses the probability of a certain transformation parameter setting $\boldsymbol{\theta}$, within the range of possible parameters, favoring more plausible transformations, with Z_{part} a normalization constant and ν a weighting factor. $\mathfrak{L}$ is an operator on the parameters $\boldsymbol{\theta}$ and represents the regularization on the transformation as $\|\mathfrak{L}(\boldsymbol{\theta})\|^2$. Regularization prevents the transformations from behaving too arbitrarily. Furthermore, regularization ensures the constrained propagation of the transformation results from areas with salient registration evidence into areas where registration evidence is largely absent or not defined. As a result, registration based, e.g., on a limited subset of floating surface points (see section 3.3.2), can be propagated to the complete set of floating surface points.

Linear transformations The simplest transformation model is the class of rigid transformations, only compensating for overall differences in pose by global translation and rotation. This rigid transformation model does not

allow any local deformation of the floating surface and is, for example, used during the patch registration problem in chapter 4. Including global scale and skew into the rigid transformation model leads to the class of affine or linear transformations, but still no local surface deformations are allowed. When using either affine or rigid transformations the regularization or parameter probablilties in (3.15) trivially reduce to a constant, because rigidity is inherent to the transformation model.

Non-linear generic transformations Using non-rigid or non-affine transformation models allows local surface deformations. Depending on the kind of transformation parametrization and regularization, different non-rigid transformation models exist from which we use two types within this thesis. A first type that we use are the generic functional transformation models obtained by a parametrization of the transformation as a weighted sum of generic basis functions similar to the variational implicit functions in (3.3) which already incorporate a mathematical regularization. The regularization typically includes a smoothness measure, ensuring local surface consistencies after transformation, and a smooth transition of the floating surface towards the target surface. The prior distribution in (3.15) will favor transformation parameters generating smoother transformations. These transformation models are mathematically well defined and are applicable to a wide range of surfaces from different objects without a learning phase and, hence, can be called generic. We use a generic transformation model with TPS basis functions for the complete facial surface registration problem in chapter 5, were a generic face surface being the floating surface is to be locally adapted and smoothly mapped towards a target facial surface from the database. Note that these generic functional models are also widely used in current computerized craniofacial reconstruction techniques.

Non-linear statistical transformations The second non-rigid transformation type that we use is an object-specific, statistical transformation model. It is constructed by a parametrization of the transformation as a weighted sum of object-specific deformations. The associated regularization is statistical in nature and represents the plausibility of the applied deformations according to an object deformation distribution. The prior distribution in (3.15) will favor transformation parameters generating more plausible object surface instances. These transformation models ensure that the surfaces are deformed in an object-specific way only, but they require a learning phase. This learning phase consists of co-registering object instances within a database to some reference object structure, using a generic non-rigid transformation model for example. Once the statistical transformation model is learned new object instances can be registered using the statistical inferred transformation model, which is shown in chapter 6.
We use a statistical non-rigid transformation model with principal component deformations learned in chapter 6 for the skull registration problem in chapter

7, were the craniofacial model template (the floating surface) is to be locally adapted and mapped in a face-specific way towards a skull surface (the target surface).

3.3.2 Similarity measure

In the context of surface registration, the similarity measure is a functional of the transformation parameters that returns a scalar number quantifying how well the floating and target surfaces are geometrically aligned (irrespective of the plausibility of the applied transformation). Similarity can be expressed using different kinds of surface characteristics, e.g. geometry (location), texture, curvature,... Using a point-based floating surface $S^f = \{\mathbf{r}_i^f | i = 1, \ldots, N^f\}$ representation, the surface similarity is typically defined over a subset $\{\mathbf{r}_j^f | i = 1, \ldots, N\}$ with $N \leq N^f$ of the floating surface points.

A general *geometry-based* individual point similarity $\mathfrak{s}_j$ implies that the point $\mathbf{r}_j^f$ is to be transformed to a new position $\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})$ at a predefined distance[4] $d_j(\boldsymbol{\theta})$ to the target surface S^t. In its most general form, this individual point similarity can be mathematically represented as a function of the transformation model parameters $\boldsymbol{\theta}$ according to:

$$\mathfrak{s}_j(\boldsymbol{\theta}) = \mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) - d_j(\boldsymbol{\theta}) \quad \text{for} \quad j = 1, \ldots, N \tag{3.16}$$

with $\mathfrak{D}(S, \mathbf{r})$ the distance between a point $\mathbf{r}$ and a surface S. Hence, optimizing this similarity measure is equivalent to eliminating the difference between the distance $\mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))$ at the currently estimated positions of the floating surface points $\mathbf{r}_j^f$ and the predefined $d_j(\boldsymbol{\theta})$ distances by finding an appropriate set of transformation parameters $\boldsymbol{\theta}$. The values of the predefined distances $d_j(\boldsymbol{\theta})$ are different for different registration problem. For the patch and the complete facial surface registration problems, the predefined distance values for points on the floating surfaces are all set to zero and independent of the transformation parameters. For the skull registration problem, on the other hand, the predefined distance values are equal to ultrasound thickness values in 52 face landmarks points of the craniofacial model template floating surface. Dependent on the craniofacial model used, the predefined distances change as a function of the transformation parameters as well.

The way the current distance $\mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))$ is calculated for a point of the floating surface to the target surface is dependent on the target surface representation. In traditional point-based or mesh-based surface representations for the target surface $S^t = \{\mathbf{r}_i^t | i = 1, \ldots, N^t\}$, the current distances are determined based on the distance between corresponding points on the two surfaces, established either manually or automatically. Given a subset of corresponding points $\{(\mathbf{r}_j^t, \mathbf{r}_j^f) | j = 1, \ldots, N\}$ with $N \leq N^f$ and $N \leq N^t$, then the

[4] In most applications, this predefined distance will be zero. However, certain applications, such as the positioning of a skin floating surface at a certain distance from a skull target surface, require the floating points be positioned at a non-zero distance from the target surface.

current distance is defined as $\|\mathbf{r}_j^t - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\|$. A typical individual point similarity $\mathfrak{s}_j$ as a function of the transformation parameters $\boldsymbol{\theta}$ using a point-based target surface representation is then defined as:

$$\mathfrak{s}_j(\boldsymbol{\theta}) = \|\mathbf{r}_j^t - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\| - d_j(\boldsymbol{\theta}) \quad \text{for} \quad j = 1, \ldots, N \tag{3.17}$$

When the predefined distances d_j are zero, then (3.17) amounts to the signed Euclidean Distance to the closest point which is a well-defined similarity criterium and the resulting transformation is a one-to-one mapping between corresponding points. However, when the predefined distances are different from zero, then (3.17) is an ill-defined similarity criterium and the resulting transformation is a one-to-many mapping. Indeed, a point can be transformed to every point on a sphere with radius $d_j \neq 0$ around the corresponding point. In order to solve this problem, additional surface information besides corresponding point positions is to be incorporated within the similarity measure favoring certain positions along the spheres of possibilities. An example is given in chapter 7 where, using a traditional craniometric point-based skull representation, the skull landmark position information is augmented with skull surface normal information favoring spherical possibilities in the direction of the positive skull surface normal.

Alternatively, we propose the use of an implicit target surface representation for the current distance $\mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))$ calculation. When using a VIS representation $\mathfrak{d}^t(\mathbf{r})$, as defined in section 3.2.2, for the target surface S^t, the current distance is straightforwardly and implicitly obtained by evaluating the transformed floating surface points $\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})$ into $\mathfrak{d}^t(\mathbf{r})$. The proposed individual point similarity $\mathfrak{s}_j$ as a function of the transformation parameters $\boldsymbol{\theta}$ using a target VIS representation is then defined as:

$$\mathfrak{s}_j(\boldsymbol{\theta}) = \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) - d_j(\boldsymbol{\theta}) \quad \text{for} \quad j = 1, \ldots, N \tag{3.18}$$

In contrast to (3.17) this similarity criterium is well-defined even with non-zero predefined distances d_j. Here, the mapping possibilities are restricted to lie on iso-distance surfaces that smoothly copy the target surface geometry at different distances from the zero-distance surface. Furthermore, no explicit notion of point correspondences is necessary, resulting in an alternative optimization strategy given in section 3.3.4, when point correspondences are not known prior to registration. But, before we discuss the optimization strategy for this kind of geometry-based VIS similarity measure, we first combine the transformation model from the previous section with the similarity definitions from this section into an objective function in the next section.

3.3.3 Objective function

The objective function combines the individual point similarity measures and the transformation model, such that the optimum (minimum or maximum) of the function corresponds to a set of transformation parameters $\boldsymbol{\theta}$ that maximizes the individual point similarities $\mathfrak{s}_j(\boldsymbol{\theta})$ while taking the transformation

regularization $\|\mathfrak{L}(\boldsymbol{\theta})\|^2$ into account. Furthermore, the objective function must be robust against noise or small errors and outliers or gross errors in the target surface data. We propose to derive this objective function using a principled statistical approach in which prior probabilities (the transformation model) are combined with likelihoods (the similarity measure) into a Maximum A-posteriori Probability (MAP) objective function. As a result, robustness against statistical noise is easily obtained, through the explicit modeling of an inlier-process. Furthermore, robustness against outliers is obtained through the explicit modeling of an outlier-process combined with the inlier-process. The final result is a better underpinned objective function quasi self-regulating robustness against noise and outliers. Furthermore, we will show equivalent closely related M-estimators from robust statistics [72] techniques for parameter estimation. Doing so, outlier influences in the objective function can be analyzed.

Noise robust MAP objective function

In this section, the registration problem is reformulated into a statistical Bayesian framework, leading to a Maximum A-posteriori Probability (MAP) objective function: *What are the most probable parameters, according to a prior transformation model, transforming the floating surface towards noisy target surface data while increasing matching quality.* Mathematically, this is similar to finding the MAP parameters maximizing the posterior probability $p(\boldsymbol{\theta}|S^t)$, which is, according to Bayes' rule, equal to:

$$p(\boldsymbol{\theta}|S^t) = \frac{p(S^t|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(S^t)} \tag{3.19}$$

resulting into the following MAP formulation, where $p(S^t)$ is omitted because it is independent of $\boldsymbol{\theta}$:

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{MAP} &= argmax_{\boldsymbol{\theta}} \left(\log p(S^t|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta})\right) \\ &= argmin_{\boldsymbol{\theta}} \left(-\log p(S^t|\boldsymbol{\theta}) - \log p(\boldsymbol{\theta})\right) \end{aligned} \tag{3.20}$$

Taking the negative log-likelihood, the MAP is restated into a minimization problem. $p(\boldsymbol{\theta})$ is the transformation model or, more generally model-likelihood, reflecting our prior belief on the possible transformations. It acts as a regularization ensuring that the deformation behaves according to our prior knowledge and was probabilistically defined in (3.15) as a Gibbs prior distribution. The first term in (3.20) is the complete data-likelihood $p(S^t|\boldsymbol{\theta})$, which expresses the probability of measuring the target surface data S^t given a transformed floating surface, using the individual point similarities defined in (3.16), while taking the target surface noise into account. The presence of noise influences the individual point similarities defined in (3.16) resulting in erroneous point similarity $\mathfrak{s}_j^*$ measurements:

$$\mathfrak{s}_j^*(\boldsymbol{\theta}) = \mathfrak{s}_j(\boldsymbol{\theta}) + \epsilon \quad \text{for} \quad j = 1, \ldots, N \tag{3.21}$$

Assuming that the errors ϵ can be modeled as additive Gaussian or normally distributed i.i.d. noise with zero mean and standard deviation σ, then the complete data-likelihood can be formulated as a product of Normally distributed evaluations with given parameters $\Theta = \{\boldsymbol{\theta}, \sigma\}$, including the unknown noise standard deviation:

$$p(S^t|\Theta) = \prod_{j=1}^{N} p(\mathfrak{s}_j|\Theta) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\mathfrak{s}_j(\boldsymbol{\theta}))^2\right) \tag{3.22}$$

$p(\mathfrak{s}_j|\Theta)$ is a a Gaussian probability density function (PDF) or distribution constituting the inlier-process, which is defined based on the point similarity measures $\mathfrak{s}_j$ centered around a zero mean with noise or error standard deviation σ. This inlier-distribution represents the probability of a residual $\mathfrak{s}_j(\boldsymbol{\theta})$ (Eq. 3.16) between a certain current distance $\mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))$ to the target surface S^t and a predefined distance d_j being explained or being an inlier according to a transformed floating surface point $\mathbf{r}_j^{f*}$ instance or realization taking into account the allowed deviation σ. Applying (3.15) and (3.22) into (3.20) gives the following *MAP objective function*:

$$\begin{aligned} \mathfrak{E}(\Theta) = &\quad \textstyle\sum_{i=j}^{N} \left[-\log p(\mathfrak{s}_j|\Theta)\right] - \log p(\boldsymbol{\theta}) \\ &= \textstyle\sum_{j=1}^{N} \left[\frac{1}{2\sigma^2}(\mathfrak{s}_j(\boldsymbol{\theta}))^2 + \log\sqrt{2\pi}\sigma\right] + \nu\|\mathfrak{L}(\boldsymbol{\theta})\|^2 + K \end{aligned} \tag{3.23}$$

With K a constant independent of Θ, which can be dropped during optimization (see section 3.3.4). The target surface noise is explicitly modeled through σ and must be estimated as well. ν is a weighing factor regulating the tradeoff between matching quality and transformation regularization, which is important when the target surface data is noisy. Choosing ν too low/high, results in an over/under-fitting of the noisy target surface data. Somewhere in between, an optimal ν value has to be chosen, fine-tuning the amount of regularization in order to generate the optimal registration result. The optimal ν value is dependent on the amount of noise present in the target surface data. To show this, the objective function is multiplied with two times the noise variance σ^2 resulting in:

$$\mathfrak{E}(\Theta) = \mathfrak{S}(\Theta) + 2\sigma^2\nu\|\mathfrak{L}(\boldsymbol{\theta})\|^2 \tag{3.24}$$

with

$$\mathfrak{S}(\Theta) = \sum_{j=1}^{N} \left[(\mathfrak{s}_j(\boldsymbol{\theta}))^2 + 2\sigma^2 \log\sqrt{2\pi}\sigma\right] \tag{3.25}$$

When setting ν equal to one, resulting in no user-specific defined amount of regularization, the tradeoff between matching quality or squared error and transformation regularization is regulated automatically (self-regulating) and solely in function of the estimated noise. However it is not obligatory to set $\nu = 1$ and to keep generality in this section we will assume $\nu \neq 1$ while working with (3.23).

The ML parameters $\Theta = \{\boldsymbol{\theta}, \sigma\}$ should satisfy $\partial \mathfrak{E}(\Theta)/\partial\Theta = 0$, minimizing the energy defined in (3.23). In most practical applications, however, the assumed distribution $p(\mathfrak{s}_j|\Theta)$ of the inlier-process is only an approximation to reality, and estimation of the parameters should not be severely affected by the presence of a limited amount of outliers. Gross errors or outliers are data severely deviating from the pattern set by the majority of the data. The evaluation of an outlier in the inlier-distribution results in a very low probability: *the outlier cannot be explained by the inlier-distribution or the outlier cannot be generated by the inlier-process.* Since

$$\lim_{p(\mathfrak{s}_j|\Theta)\to 0} \log p(\mathfrak{s}_j|\Theta) = -\infty \tag{3.26}$$

the contribution to the data log-likelihood of an observation that is a-typical for the inlier-distribution is very high. Dealing with this requires the application of robust statistics, which provides tools for problems in which underlying assumptions (inlier-distribution) are inexact. A robust procedure should be insensitive to departures from underlying assumptions caused by, for example, outliers. That is, it should perform well under the underlying assumptions and the performance should deteriorate gracefully as the situation departs from the assumptions.

In appendix A we present two main approaches for robust statistics and apply it to the surface registration problem. The first approach makes use of robust estimators that are substituted for the classical estimators such as the sum of squared errors (when the residuals are normally distributed). It is, however, unclear how the analytical form of these estimators as well as their internal parameters have to be chosen upfront. The second approach models the outliers explicitly as well, which allows the modeling of additional constraints such as, e.g. spatial coherence of outliers. Furthermore, it can be shown that for every such outlier process, an equivalent robust estimator from the first approach can be constructed. While this approach leads to a robust estimator of a general functional form, it still requires the ad hoc setting of certain parameters.

In the next section, following Fransens *et al.* [73], we will present a principled approach to link the robust estimators to a statistical model of how the measurement, including the outliers, were generated. Indeed, the outlier-process will be modeled as a statistical process as well, akin to the inlier-process. As a result, making use of the theory expanded in appendix A, we will be able to construct equivalent M-estimators for outlier rejection. The final result will be an objective function with quasi self-regulating outlier robustness, depending on the kind of outliers, besides self-regulating noise robustness obtained previously.

Outlier robust MAP objective function

We will model both the inliers and outliers as random variables within the complete data log-likelihood resulting in a robust copy of the first term in

(3.23). The transformation model-likelihood or the second term in (3.23), remains unchanged. Therefore, we will concentrate on the complete data log-likelihood term from now on.

Suppose a data sample $\mathfrak{s}_j$ is generated by either an inlier-process or an outlier-process, then the generating complete-process can be specified by conditioning the sample likelihood on the state of a binary-valued latent variable z_j, signalling whether a data sample $\mathfrak{s}_j$ was generated by the inlier-process $z_j = 1$ or not $z_j = 0$:

$$p(\mathfrak{s}_j|z_j,\Theta) = \begin{cases} p(\mathfrak{s}_j|\Theta) & \text{if} \quad z_j = 1 \\ p^o(\mathfrak{s}_j) & \text{if} \quad z_j = 0 \end{cases}$$
$$p(\mathfrak{s}_j|z_j,\Theta) = p(\mathfrak{s}_j|\Theta)^{z_j} p^o(\mathfrak{s}_j)^{(1-z_j)} \text{ for } z_j \in \{0,1\} \tag{3.27}$$

$p(\mathfrak{s}_j|\boldsymbol{\theta})$ is a Gaussian[5] probability density function (PDF) constituting the inlier-process and was defined in (3.22). $p^o(\mathfrak{s}_j)$ is the outlier PDF constituting the outlier-process for which several choices can be considered. First off all, if some knowledge about the outlier generating process is known, then $p^o(\mathfrak{s}_j)$ can be set to a known distribution. In most cases, however, no outlier generating process knowledge is given and $p^o(\mathfrak{s}_j)$ can either be estimated similar to the inlier-distribution or can be set to a fixed uniform distribution for example. The latter is the case for our surface registration problems. Therefore, we take a fixed outlier distribution, which is assumed to be uniformly distributed $p^o(\mathfrak{s}_j) = \delta$. Spatial constraints on the outliers are typically incorporated using a proper modeling of z_j depending on the prior knowledge of the outlier problem besides the outlier PDF.

In appendix B we derive the objective functions for robust parameter estimation in the presence of outliers. Depending on the type of outlier generating process and thus the type of latent variable (deterministic or random, or a combination of both), different functionals are derived. Here we summarize the major results for the different types of latent variables.

Random Bernoulli LV The most general type of latent variable we consider is a random (probabilistic) variable, where the binary latent variable z_j'' is Bernoulli distributed and the outliers are uniformly distributed ($p^o(\mathfrak{s}_j) = \delta$). The consequence is that the outlier map $\mathbf{z}''$ is a random map with an associated prior-distribution $p(\mathbf{z}'')$. Let P be the prior probability of z_j'' being an inlier (i.e the fraction of observations $\mathfrak{s}_j$ thought to be generated by the inlier-process) and let $P_o = 1 - P$ be the prior probability of being an outlier (i.e the fraction of observations $\mathfrak{s}_j$ thought to be generated by the outlier-process).

[5] In general, $p(\mathfrak{s}_j|\boldsymbol{\theta})$ can be any probability density function, but in our application, we have assumed the measurement noise to be Gaussian, hence the choice of this particular PDF.

Then, assuming $\mathbf{z}''$ is i.i.d[6], $p(\mathbf{z}'')$ can be specified as follows:

$$p(\mathbf{z}'') = \prod_{j}^{N} P^{z''_j} P_o^{(1-z''_j)} \tag{3.28}$$

This kind of random latent variable choice is used for the skull registration problem. Outliers within this registration problem are due to gross errors in the skull acquisition process and the construction and choice of the skull representation. These errors are treated probabilistically. Further details are given in chapter 7.

Deterministic LV As a special case, we consider a binary latent variable z'_j, with very strong prior knowledge encoded in a deterministic (non-probabilistic) outlier detection function f that directly determines wether an observation $\mathfrak{s}_j$ is either an inlier or outlier:

$$z'_j = f(\mathfrak{s}_j(\boldsymbol{\theta})) \in \{0, 1\} \tag{3.29}$$

This kind of deterministic latent variable choice is used for the patch registration and complete facial surface registration problems. Outliers within these registration problems are due to partial overlap between the target and floating surfaces. Individual point similarities defined in section 3.3.2 measured in floating surface points r_j^f not belonging to the partial overlap are very erroneous and therefore outliers. The deterministic function f generating binary valued latent variables can be defined based on the VBF $\mathfrak{b}(\mathbf{r})$ of the target surface. Further details are given in chapters 4 and 5.

Combined deterministic and random LV The final choice of latent variable we consider is not really different from the two previous, but rather a combination of them. Imagine an incomplete skull substrate is given for skull registration or an incomplete patch, or incomplete facial surfaces. Suppose furthermore that besides these incomplete data, probabilistic outliers are present as well for these surfaces. We thus have combination of the two types of latent variables: deterministic, coding for missing data, and random, coding for the presence of erroneous data. To tackle these problems we define a combined latent variable outlier map $\mathbf{z} = \mathbf{z}' \wedge \mathbf{z}''$ with $\wedge$ a conjunction and $(1 - \mathbf{z}) = (1 - \mathbf{z}') \vee (1 - \mathbf{z}'')$ with $\vee$ a disjunction incorporating a regular deterministic latent variable $\mathbf{z}'$ outlier map and a random latent variable $\mathbf{z}''$ outlier map according to:

$$\mathbf{z}' = f(\mathbf{x}) \tag{3.30}$$

and

$$p(\mathbf{z}'') = \prod_{j}^{N} P^{z''_j} P_o^{(1-z''_j)} \tag{3.31}$$

[6]i.i.d. = independent and identically distributed

Using these latent variables, the first term in our MAP-objective function (3.23) (complete data negative log-likelihood), can then be reconstructed using the complete-process $p(\mathfrak{s}_j|z_j,\boldsymbol{\theta})$ defined in (3.27) and taking the negative logarithm:

$$\mathfrak{S}(\Theta,\mathbf{z}',\mathbf{z}'') = \sum_{j=1}^{N} -[z_j' z_j''] \log p(\mathfrak{s}_j|\Theta) - [(1-z_j') + z_j'(1-z_j'')] \log \delta \tag{3.32}$$

Using the procedures defined in appendix B, we can estimate and/or eliminate the latent variables from this equation, resulting in the following final functional:

$$\mathfrak{S}(\Theta) \quad = \sum_{\mathbf{z}''} \left[\sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(f(\mathfrak{s}_j(\boldsymbol{\theta})), z_j''), \beta(f(\mathfrak{s}_j(\boldsymbol{\theta})), z_j'')) \right] \tag{3.33}$$

where

$$h(x, \alpha(z',z''), \beta(z',z'')) = \alpha(z',z'')x^2 + \beta(z',z'') \tag{3.34}$$

with

$$\begin{aligned} \alpha(z',z'') &= \tfrac{1}{2\sigma^2}[z'z''] \\ \beta(z',z'') &= -[(1-z') + z'(1-z'')] \log \delta + [z'z''] \log \sqrt{2\pi}\sigma \end{aligned} \tag{3.35}$$

Following an out-integration strategy for the random latent variable we can finally define the equivalent M-estimator as:

$$\rho(x,\alpha,\beta) = h(x, \alpha(f(x),b), \beta(f(x),b)) \tag{3.36}$$

To summarize, we obtain a general complete MAP objective function from which we will start to solve the different registration problems given in the relevant chapters:

$$\mathfrak{E}(\Theta,\mathbf{z}',\mathbf{z}'') = \mathfrak{S}(\Theta,\mathbf{z}',\mathbf{z}'') + \nu \|\mathfrak{L}(\boldsymbol{\theta})\|^2 \tag{3.37}$$

such that after latent variable estimation the objective function becomes:

$$\mathfrak{E}(\Theta) = \mathfrak{S}(\Theta) + \nu \|\mathfrak{L}(\boldsymbol{\theta})\|^2 \tag{3.38}$$

with

$$\mathfrak{S}(\Theta) = \sum_{j=1}^{N} \rho(\mathfrak{s}_j(\theta), \alpha, \beta) \tag{3.39}$$

where

$$\rho(\mathfrak{s}_j(\theta),\alpha,\beta) = h(\mathfrak{s}_j(\theta), \alpha(f(\mathfrak{s}_j(\theta)), b_j), \beta(f(\mathfrak{s}_j(\theta)), b_j)) \tag{3.40}$$

In the next section we give details about possible optimization strategies dependent on the target surface representation w.r.t the transformation parameters $\boldsymbol{\theta}$ and the noise standard deviation σ given the equivalent M-estimator formulation in (3.38) based on (3.37) after latent variable estimation.

3.3.4 Optimization

The optimization method defines how the parameters within the objective function are adjusted or updated to lower[7] the objective function evaluation and therefore improving the matching quality while taking into account the transformation model regularization. Optimization is a broad discipline in mathematics where a lot of methods have been proposed with varying complexity. In the previous section we already saw that for an objective function to be robust against outliers with prior knowledge a joint parameter optimization or estimation is necessary (3.37). Furthermore, according to [74, 75, 76, 77] this joint optimization can be solved by alternating between a latent variable optimization and a ML parameter optimization while during either optimization or update the other update or estimate is kept fixed. We showed that the latent variable optimization could be performed using minimization by determination of the infimum (this is in fact a combinatorial optimization), deterministic function evaluation (not really considered an optimization) and out-integration (Mean-Field optimization) dependent on the kind of outlier latent variable. Note that the latter case where the latent variable is a random probabilistic variable using out-integration is in fact the Expectation step (E-step) within an Expectation-Maximization (EM) optimization algorithm, which has desirable convergence properties. Therefore the ML parameter estimation or optimization is considered to be the Maximization step (M-step), which was reformulated into a minimization by taking the negative logarithms in (3.20), within the same EM optimization. Using a broad interpretation of the E-step (Latent Variable step: LV-step) making it applicable to non-probabilistic outlier latent variables as well, an EM optimization produces a sequence of parameter estimates $\{\widehat{\Theta}^{(m)}|m = 0, 1, \ldots\}$ by alternating the following two steps:

LV-step or E-step: On the $(m+1)^{th}$ iteration, the expectation of the complete data log-likelihood $\mathfrak{S}(\Theta, \mathbf{z}', \mathbf{z}'')$ is obtained by determining the latent variable values based on the ML parameter estimation $\widehat{\Theta}^{(m)}$ with $\Theta = \{\boldsymbol{\theta}, \sigma\}$ of the previous iteration. The expectation of the complete log data-likelihood or the so-called Q-function is then given by:

$$Q(\Theta|\widehat{\Theta}^{(m)}) = -\mathfrak{S}(\Theta) \tag{3.41}$$

with $\mathfrak{S}(\Theta)$ defined in (3.39). The LV-step is in fact the determination of the equivalent M-estimator used in the following M-step.

M-step In the case of MAP estimation, the ML parameters are then updated according to:

$$\widehat{\Theta}^{(m+1)} = \text{argmin}_{\Theta}\{-Q(\Theta|\widehat{\Theta}^{(m)}) - \log p(\theta)\} \tag{3.42}$$

[7] We assume that the objective function is always written such that the optimum corresponds to a minimum of the objective function

leading to the following objective function or energy minimization:

$$\mathfrak{E}(\Theta) = \mathfrak{S}(\Theta) + \nu \|\mathfrak{L}(\boldsymbol{\theta})\|^2 \tag{3.43}$$

which is the same as in (3.38). The E-step and the M-step are alternated during each iteration until convergence or when minimal changes in parameter updates occur. In this section we will concentrate on the M-step of the joint EM optimization which is the same for every choice of outlier latent variable but can be different for a mesh or point-based target surface representation and an implicit target surface representation.

A prerequisite for optimization is to compute the individual point similarities $s_j(\boldsymbol{\theta})$ defined in (3.17) and (3.18) for the point-based and implicit target surface representation respectively. For the proposed implicit target surface representation this comes down to simple VIS evaluations. Due to the smooth and analytical properties of these VIS, gradient information can always be calculated and the ML parameter update can be obtained according to:

$$\Theta^{(m+1)} = \Theta^{(m)} - s\nabla\mathfrak{E}(\Theta) \tag{3.44}$$

This is a Steepest-Descent optimization approach including knowledge of the gradient. The parameters are updated by taking a step s in the negative direction of the gradient, which is calculated analytically $\nabla\mathfrak{E}(\Theta) = [\partial\mathfrak{E}/\partial\Theta_1, \ldots, \partial\mathfrak{E}/\partial\Theta_n]$ based on the first derivatives of (3.38) to every parameter in Θ. The optimal step s is obtained using a Brent's line minimization [78]. Inclusion of the first derivatives increases robustness and reduces the number of iterations. More advanced Steepest-Descent methods can be applied as well, like Conjugate gradient (used for patch registration) and quasi Newton (used for skull registration). These will gradually build up and use the Hessian or second order derivatives for faster convergence. Furthermore, Stochastic Steepest-Descent methods can also be applied to decrease computational efforts and to increase robustness when the amount of floating surface points is huge. It must be noted that the term $-[(1-z') + z'(1-z'')]\log\delta$ in $\beta(z', z'')$, originating from the outlier-process, after the LV-step and used during the M-step is independent of the ML parameters Θ and can be treated as a constant during first derivative calculations, therefore vanishing.

For point-based target surface representations corresponding points $\{(\mathbf{r}_j^t, \mathbf{r}_j^f) \mid j = 1, \ldots, N\}$ are required to compute the individual point similarities $s_j(\boldsymbol{\theta})$. Therefore an additional correspondence search is needed within the M-step before a ML parameter update can be performed. These searches are computational expensive and require extensive implementation tricks (e.g. Octree representations) to reduce the computational cost of finding corresponding points. A corresponding point $\mathbf{r}_j^t$ for a floating surface point $\mathbf{r}_j^f$ is typically defined being the closest point to $\mathbf{r}_j^f$. It must be noted that the quality of the correspondences based on a non-continuous mesh target surface representation is dependent on the surface point sampling leading to discretisation errors within the correspondences. When the predefining distances d_j are zero, an exact one-to-one mapping between corresponding points is defined

and an update of the ML-parameters can be obtained using the well known Iterative Closest Point (ICP) optimization [59]. The correspondence search is alternated with the transformation parameter update in order to improve correspondences and transformation estimates. These transformations are obtained by solving a closed (meaning not strictly requiring an iterative scheme) problem such that corresponding points are mapped onto each other:

$$\mathbf{r}_j^t = \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) \tag{3.45}$$

Of course individual point confidences (inlier-beliefs) and the transformation regularization are to be taken into account while solving the closed problem. Alternatively and especially when the predefined distances are not equal to zero $d_j \neq 0$ the Steepest-Descent in (3.44) can be applied as well once the correspondences are established.

Note that correspondences being closest points can be made explicit using an implicit sDT target surface representation according to the following implicit to explicit conversion:

$$\mathbf{r}_j^t = \mathbf{r}_j^f - \mathfrak{d}^t(\mathbf{r}_j^f)\frac{\nabla\mathfrak{d}^t(\mathbf{r}_j^f)}{\|\nabla\mathfrak{d}^t(\mathbf{r}_j^f)\|} \tag{3.46}$$

Using an implicit sDT target surface a corresponding point $\mathbf{r}_j^t$ of $\mathbf{r}_j^f$ is obtained by walking an amount equal to the VIS evaluation or signed distance value $\mathfrak{d}^t(\mathbf{r}_j^f)$ in the negative direction of the normalized VIS gradient $\nabla\mathfrak{d}^t(\mathbf{r}_j^f)/\|\nabla\mathfrak{d}^t(\mathbf{r}_j^f)\|$. Doing so, the ICP optimization strategy (3.45) can applied on implicit target surfaces as well. Furthermore due to the continuous nature of the VIS, no discretisation errors are present in the established point correspondences. Whether to use the ICP strategy (3.45) or the Steepest-Descent strategy (3.44) depends on the transformation model and the complexity or feasibility to obtain a closed problem formulation. When the amount of parameters is limited or when it is not feasible to obtain a closed form formulation (e.g. no one-to-one mapping), it is preferred to apply the Steepest-Descent strategy. This is for instance the case for the patch and skull registration problems. When the amount of transformation parameters is equal to the amount of points on the floating surface and if a closed problem formulation can be obtained, it is interesting to apply the ICP strategy. This is the case with the complete facial surface registration using TPS non-rigid transformation models, incorporating a vast amount of parameters. Furthermore, using the TPS transformations an elegant closed form formulation, taking into account the transformation regularization, can be obtained and applied. Therefore, an ICP-like strategy is used in chapter 5 for the facial surface registration problem in which correspondences are obtained according to (3.46).

A final issue concerning optimization is the fact that the objective function is not guaranteed to be convex. This means that the optimization procedure may converge to or get stuck in a local minima instead of the global correct minimum. The original Least-Square formulation using the Quadratic estimator in (A.2) is convex and does not suffer from local minima, but is not

robust against outliers. The robust outlier process formulation, however, may not be convex. Furthermore, the non-convexity or the amount of local minima increases when having a vast amount of transformation parameters to estimate and when having erroneous or conflicting point-correspondences. Fortunately, converting outlier processes in terms of robust estimators (E-step) results in the possible application of deterministic continuation methods to tackle the local minima problem. One such method is deterministic annealing [79] and is used for the facial surface registration problem. The idea is to choose a control parameter within the estimator that can be used to change the shape of the estimator to construct a convex approximation of the objective function. The control parameter is slowly adjusted such that the objective function increasingly approximates the original non-convex estimation.

Further optimization details are given in the relevant chapters about patch registration (chapter 4), facial surface registration (chapter 5) and skull registration (chapter 7).

3.3.5 Initialization

The proposed surface registration method belongs to a class of registration algorithms in which an objective function, reflecting the quality of registration, is iteratively optimized. Consequently, these methods can generate accurate results. However due to their iterative nature they need to be initialized with a proper estimate of the unknown pose parameters. Prior to registration, the floating surface is to be brought into the vicinity of the target surface, by repositioning the floating surface, using a rigid or affine transformation, into the coordinate system of the target surface. Furthermore, final registration results are dependent on the quality of the initial pose estimate. The latter is to be within the convergence range of the iterative registration method in order to generate successful registration results. Due to their dependency on the initial pose, registration algorithms iteratively optimizing an objective function are refereed to as *constrained registration* methods and due to their accuracy they are also referred to as *fine registration* methods.

Finding relative poses with arbitrary differences in initial pose is referred to as *unconstrained registration* or *crude registration*. Typically a coarse-to-fine strategy is applied. First, a crude registration is performed to give an initial estimate of the pose. Next, a fine registration is applied to compute an accurate transformation. Consequently, crude registration is not to be accurate but is successful as long as it gives a correct initialization for the fine registration. Having a more accurate crude registration does have the advantage that the fine registration can converge more robustly and within fewer iteration steps.

Initialization or crude registration is often solved manually. A first way is to interactively rotate and translate the surfaces manually to bring them into each others proximity. An alternative is to manually indicate a sparse set of prominent point correspondences based on which the initial pose transformation parameters are computed according to (3.45). The latter is done for the complete facial surface and skull registration problems both using the

52 face landmarks as prominent point correspondences to determine an initial translation and rotation. When the amount of surfaces to register increases as in the patch registration problem, it becomes interesting to have an automatic initialization or crude registration method. We developed such an automatic initialization method for the patch registration problem based on VIS-Spin images (VISS-images) to find prominent point correspondences without manual user intervention. Further details thereabout are given in chapter (4).

3.4 Conclusion

In this chapter, a general noise- and outlier-robust surface registration methodology, using implicit target surface representations, was presented to solve three different registration problems required as part of our statistical craniofacial modeling and reconstruction procedure. In section 3.2, we presented an implicit surface representation commonly used throughout the thesis to represent the target surfaces. In addition, a variational boundary function that retains the boundary concept of open surfaces within an implicit surface representation, was created. The thus constructed implicit surfaces are an approximation of a signed Distance Transform, incorporating interesting implicit notions of distance and direction for surface registration purposes. Using these target surface representations, we developed a general robust surface registration framework in section 3.3. Components generally needed for registration were introduced while focussing on the use of implicit surfaces and a robust statistical objective function. The latter was obtained by reformulating the objective function into a robust MAP problem, with explicit inlier- and outlier-process modeling. Using an explicit outlier-process formulation was done to incorporate prior outlier knowledge within the objective function leading to a joint optimization. Furthermore, equivalent M-estimators (as used in robust statistics) were constructed for different kinds of latent variables (which code for the presence of an outlier). Doing, so we were able to analyze outlier influences. The construction of the M-estimators was obtained during the E-step of the joint estimation, regarded as a general interpretable EM optimization. Thanks to the implicit surface representations, no additional correspondence search is needed and a straightforward Steepest-Descent M-step optimization could be applied. Furthermore, well established optimization routines for robust statistics are directly applicable on the explicit outlier-process formulations due to the unification with robust M-estimators.

The final result is a unifying theoretical surface registration methodology, which is extensively used throughout the thesis for almost every component of our a statistical craniofacial reconstruction technique. Starting from the theoretical surface registration foundations given in this chapter, we solve every registration problem that is encountered. Specific details about the different registration problems are given in chapter 4, 5 and 7 for the patch, complete facial surface and skull registration problems respectively.

Part II

Facial database acquisition and registration

Chapter 4

Facial surface acquisition

4.1 Introduction

Besides soft-tissue thickness acquisition (cf. appendix C), the second prerequisite for a craniofacial database is the acquisition of facial surface information preferably from living subjects. In chapter 2 several possibilities used in practise were enumerated (e.g. CT, MRI, laser and photogrammetric scanners) all with their advantages and disadvantages. To acquire face information (geometry and texture) of living subjects in an upright position a non-invasive mobile photogrammetric scanner (ShapeCam, Eyetronics (www.eyetronics.com)) is used, which is depicted in figure 4.1. This is an active 3D capturing device, which projects a regular grid on the face (figure 4.1(middle)). At the same time a digital camera (CANON D60) takes an image of the face from a different point of view such that by triangulation the 3D shape of the face can be retrieved as a dens set of connected 3D points. The ShapeCam also captures a second image incorporating texture information, which is used to determine the 3D location of the landmarks on the skin surface reconstruction. Indeed, before taking 3D images, the 52 landmarks are marked on the face with a blue eyeliner pencil by a forensic odontologist. The coordinates of these blue points are extracted by simple image processing. Afterwards the colored texture information is converted into gray-value texture information for recognition purposes (cf. chapter 1 and 2).

An advantage of the ShapeCam is its non-invasiveness and portability, similar to the Epoch 4B, making it easier to create a database over a sufficiently large and diverse population. Furthermore, faces of subjects can be acquired in an upright position. A disadvantage of the ShapeCam, which is common for similar 3D acquisition devices, is the limited viewing-angle of the camera, resulting in a partial surface acquisition (figure 4.1(right)) of the complete facial surface. The reason is that the acquisition system can only compute 3D information of what can be "seen"from a certain viewpoint. Therefore, a complete three-dimensional facial surface is assembled from several partial surface acquisitions from unknown viewpoints (due to the camera portability).

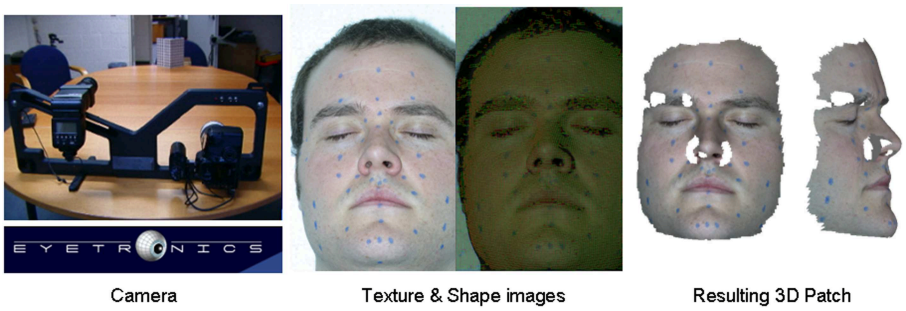


Figure 4.1: Illustration of the Eyetronics camera (left). Per viewpoint two images are taken: a regular texture image and an image with a projected grid for shape or geometry extraction (middle). The result per viewpoint is a patch or partial surface consisting of 3D shape information combined with texture information (right).

The 3D acquisition from a single viewpoint is referred to as a patch or partial surface. A complete surface is build from many patches or partial surfaces and can be seen as a patchwork. The number of views necessary is mainly determined by the required detail, camera resolution and the complexity of the object. Furthermore, in the absence of any calibration data relating the different views (as with free-hand cameras), tedious manual alignment is often used to initiate further computer-based refinement. As a result, the construction of complex 3D objects, or large databases of 3D objects, cannot be practically implemented unless a fully automatic reconstruction system is available.

Combining several patches into a single surface involves two main phases (which was already shown in figure 1.2). First the patches need to be aligned accurately into a common coordinate frame, the *registration phase*. Secondly, the registered patches need to be integrated into a single entity, which is the *integration phase*. Furthermore, both the registration and the integration must be robust against noise and outliers in the patches due to the acquisition process and due to the partial overlap in between different patches. In order to have a fully automatic assembly, without any kind of a priori knowledge of the geometry of the imaging process or user interaction, more than one registration task ir required. First, an initial alignment needs to be calculated by means of a crude registration (cf. [80], e.g.), followed by a pose refinement using a fine registration (cf. [59], e.g.). The former is a global registration problem or initialization where no a priori information on relative positions and orientations of the patches is available (figure 4.2). The latter or pose refinement assumes that an initial, but inaccurate, estimate of the exact positions is known (figure 4.6). When more than two patches are to be registered, both crude and fine pair-wise registrations need to be combined into a multi-view registration (figures 4.4 and 4.7). Finally, the registered patches need to be integrated into one entity or surface (figure 4.8). Instead of selecting the

best algorithm available for every subtask, it would be interesting to have a single common representation of all patches throughout the whole pipeline of crude- and fine registrations, even up to the integration stage, like in [81, 82].

In order to provide a fully automatic, robust and accurate method for aligning and integrating partial reconstructions without any prior knowledge of the relative viewpoints of the camera or the geometry of the imaging process, we developed a 4-step registration and integration algorithm based on common VIS representations for open surfaces (cf. chapter 3) of the patches. This implicit signed Distance representation for patches has several useful properties for solving the different registration and integration tasks and doesn't require a memory inefficient and resolution limiting discrete grid. Furthermore, additional smoothing properties of the VIS representation [69] allow us to deal with noisy patch data. In section 4.2 we concentrate on the registration phase, while in section 4.3 the focus is on the integration phase. Experimental results on real-live data, noiseless - and noisy simulated data are given in section 4.4 to show the practicality, the accuracy and robustness of the developed automatic patch registration and integration algorithm for facial surface acquisition. A final conclusion is given in section 4.5.

4.2 Patch registration framework

Our complete automatic registration algorithm for aligning multiple partially overlapping surfaces or surface patches consists of a coarse-to-fine strategy as in [80] resulting into four stages. First, a robust but less accurate pair-wise crude registration between every pair of patches is calculated in subsection 4.2.1 based on point correspondences, which are established by comparing directed point-signatures calculated from the VIS representations. A pose clustering procedure [83] selects the best rigid transformation from valid transformations that map corresponding point pairs, while making their directed (normal) components coincide. Local 2-D geometry histograms calculated from the VIS representations are used as point-signatures, which are an improved variant of the spin-images in [84]. Second, a selection of all pair-wise crude registrations are combined into a multi-view crude registration in subsection 4.2.2 by using a minimum spanning tree algorithm (MST) like [80]. We define the pair-wise transformation cost in terms of the number of inliers supported by the transformation. The result of the MST algorithm is a non-redundant set of relative pair-wise transformations that transform all patches in the same coordinate system. This set of initial transformations is further improved, in a third step (subsection 4.2.3), by applying pair-wise fine registrations, which are less robust in terms of initialization but more accurate. These fine registrations are calculated iteratively by a gradient descent minimization of a distance error criterion, based on the continuous VIS representation. In order to deal with partial overlap between patches, a novel deterministic outlier detection function is defined, which is independent of the relative pose or distance histograms between two partial surfaces. Both the minimization and outlier handling are

defined and combined using the theoretical robust surface registration framework from chapter 3. Finally, in a fourth step elaborated in section 4.2.4, we cope with pair-wise registration error accumulation by applying a simultaneous multi-view fine registration of all partial surfaces with intermediate integrations. The fine registration of a patch with an intermediate integrated surface is similar to the pair-wise fine registration. A common representation is used throughout the whole registration pipeline. The four stages will be elaborated in the subsequent sections, while shortly introducing related issues and alternatives. For an exhaustive and more elaborated related work review, we refer the reader to Appendix D.

4.2.1 Pair-wise crude initialization

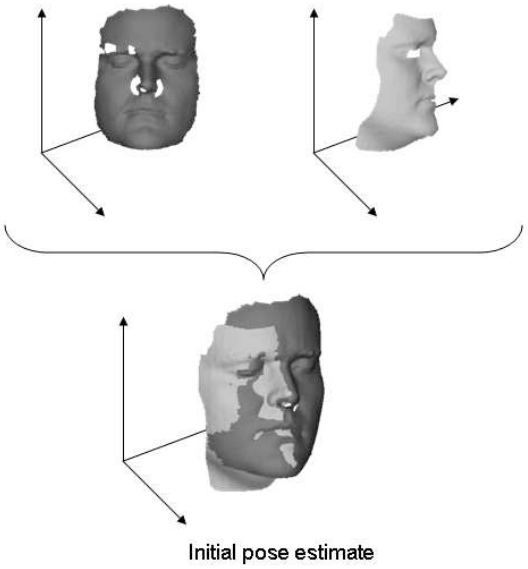


Figure 4.2: The purpose of pair-wise crude registration: Given two patches within their coordinate systems, bring them both into a common coordinate system by applying an initial pose estimate.

The first step in our registration pipeline is the crude alignment of patch pairs starting from an arbitrary relative pose. The purpose of pair-wise crude registration or initialization is depicted in figure 4.2. Given a patch being the target surface and a patch being the floating surface both in their own coordinate system without any prior positioning knowledge, the goal is to reposition the floating surface within the coordinate system of the target surface, such that overlapping parts of the surfaces initially align. Crude registration is considered successful if it results in a proper initialization for a following fine

registration. Having a more accurate crude registration does have the advantage that the fine registration can converge more robustly and within fewer iteration steps.

Crude registration is often solved manually or with a priori knowledge about the geometry of the imaging process. The former is done by letting a user indicate point correspondences or rotate and translate the patches manually. This is a common approach in many commercial scanner systems. The latter is done mechanically or based on calibration information of the used scanner. Of course these are scanner dependent and sometimes additional mechanical tracking hardware is required. Two different approaches for automatic crude registration without knowledge about the 3D sensor exist: a *pose search* approach and a *correspondence search* approach. In a pose search approach a vast amount of pose estimates to initialize a pose refinement algorithm are tested and the best result after fine registration is retained. Correspondence search approaches are computationally more interesting and try to establish point correspondences between the floating and the target surfaces by constructing and comparing point-descriptors (special features or point-signatures). Based on these correspondences a transformation estimate is determined in a robust way, in order to deal with possible false correspondences.

Our robust pose estimation is a correspondence search approach based on directed point-signatures, which are descriptors of the local geometry around a point $\mathbf{r}$ on a surface, in combination with a robust voting scheme similar to [83]. The directed component of the point-signature is the normal $\mathbf{n}$ in the point $\mathbf{r}$ to the patch surface. The point-signature is an improved variant of the spin-images defined by Johnson [84].

A spin-image is a mapping of the local 3D geometry around a point $\mathbf{r}$ to a 2D histogram. To create a spin-image of a point $\mathbf{r}$, a plane with parameters (u, v) through $\mathbf{r}$, parallel to the normal $\mathbf{n}$, is created. Along the vertical axis of the spin-image the distance v_i of every point $\mathbf{r'}_i$ in the neighborhood of $\mathbf{r}$ to the tangent plane at point $\mathbf{r}$ is plotted, while along the horizontal axis the distance u_i between $\mathbf{r}$ and $\mathbf{r'}_i$ projected on the tangent plane according to the normal in $\mathbf{r}$ is plotted. Based on the spin-map coordinates, which are similar to a local cylindrical mapping,

$$\begin{aligned} u_i &= \sqrt{\|\mathbf{r} - \mathbf{r'}_i\|^2 - (\mathbf{n} \cdot (\mathbf{r} - \mathbf{r'}_i))^2} \\ v_i &= \mathbf{n} \cdot (\mathbf{r} - \mathbf{r'}_i) \end{aligned} \tag{4.1}$$

one or more bins $(u, v)_i$ in the spin-image are incremented for every point $\mathbf{r'}$. The plane is then rotated (spinned) around the normal $\mathbf{n}$ for a finite number of angles θ, hence the name spin-image. An important advantage of the spin-images is that they produce a local shape description of the patch surface, which is invariant under rigid motion. However, traditional spin-images have several shortcomings. First of all, they are created based on a mesh representation of the patches and as such are dependent on the surface sampling. Thanks to a bilinear interpolation the influence of a vertex for a particular bin is distributed over the (four) surrounding bins, but in order

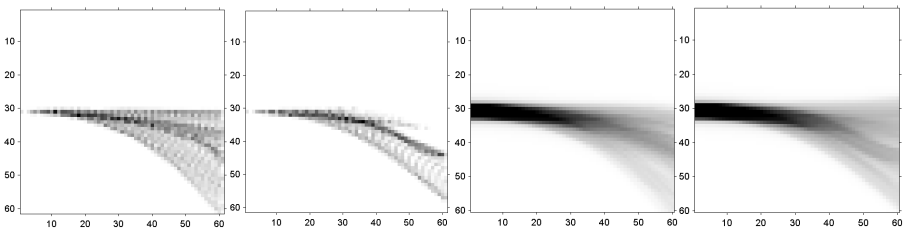
to create consistent spin-images at least a uniform mesh surface sampling is necessary. Secondly, a point laying at the outer border of a patch or at the border of a hole in the patch will generate an incomplete spin-image, while in another patch the same corresponding point may not lay at a border and generate a complete spin-image. Both spin-images of the same point in different patches will show small resemblance. Thirdly, determination of neighboring points of every point $\mathbf{r}$ for which a spin-image is created, together with an additional bilinear interpolation is computational expensive. Finally, [84] proves, by experiment, that the sampling frequency between two different patches may differ only at the condition that the bins are big enough. However, bigger bins lead to a lower resolution and a lower descriptive power of the spin-images. The bottom line is that working with mesh-based or traditional spin-images requires spin-image parameter fine-tuning towards the mesh tessellation and therefore requires some expertise.

To overcome the shortcomings of mesh-based spin-images we propose a new way to calculate spin-images, based on the VIS representation of the patches, which we call *Variational Implicit Surface Spin-images* or *VISS-images*, as follows. A plane Im through a point $\mathbf{r}$ of a patch and its normal $\mathbf{n}$ is defined and is equidistantly sampled at points $\{\mathbf{b}_i |\ \mathbf{b}_i \in Im\}$ corresponding to the centers of the bins (u, v) in Im. Every sampling point of the plane is evaluated in the VIS $\mathfrak{d}(\mathbf{r})$ of the patch (cf. section 3.2.3) and the corresponding bin $(u, v)_i$ is then incremented with a value inversely proportional to the function value $\mathfrak{d}(\mathbf{b}_i)$, for instance:

$$(u, v)_i = (u, v)_i + \frac{1}{max(1, |\mathfrak{d}(\mathbf{b}_i)|^\gamma)} \tag{4.2}$$

with $\gamma \geq 1$. The plane is then rotated around the normal in $\mathbf{r}$ over an angle θ and every point associated to each bin center is evaluated again and the bin values are updated. This is done over angles $\theta_i = i2\pi/s$ with $i = 1, \ldots, s-1$ where s determines the accuracy of the spin-images. Plane points $\mathbf{b}_i$ close to the patch will have a larger entry in the spin-image than points further away. A larger value of γ will generate a more accurate VISS-image at the condition that s is high. However, the VISS-image is then dependent on the starting position of the plane. A lower value of γ will generate a smoother VISS-image, less dependent on the starting position of the plane Im, due to the decreased suppression of plane bin center points $\mathbf{b}_i$ further away from the patch surface.

VISS-images, compared to the classic spin-images, are independent of the patch surface sampling and thanks to the interpolation and extrapolation properties of the VIS representation, they are more reliable at borders and holes. A VIS representation only needs to be calculated once and the images are generated based on simple function evaluations, which is more efficient and easier than determining spin-map coordinates for every point in the neighborhood of a point $\mathbf{r}$. Furthermore, no extra bilinear interpolation is necessary. Figure 4.3 depicts two spin- and VISS-images of a corresponding point-pair in between a target and a floating patch with partial overlap. The point on the floating patch is situated in the vicinity of the patch border. It is



(a) (b) (c) (d)

Figure 4.3: Comparison between spin-images (a,b) and VISS-images (c,d) of a corresponding point pair on different patches , where the point of the second patch lies in the vicinity of a border (b,d).

clearly observed that the corresponding spin-image (figure 4.3(b)) is incomplete, therefore showing small resemblance with the spin-image of the point on the target patch (figure 4.3(a)). In contrast, the resemblance between the two VISS-images (figure 4.3(c) and (d)) is preserved, due to the extrapolation of the border. Furthermore, it is seen that the VISS-images are smoother than the spin-images. Values in neighboring bins of VISS-images are more alike and the image clearly gives a better continuous approximation of a 2D geometry histogram. Spin-images are much more dependent on the number of vertices in a mesh and whether a vertex falls within a bin or not, despite the use of bilinear interpolation.

Corresponding point-pairs between a floating and a target patch are found by calculating VISS-images for a large number of randomly selected points in both patches and by calculating a similarity measure, for instance, the normalized correlation coefficient, between them.

$$corr(Im_1, Im_2) = \frac{cov(Im_1, Im_2)}{\sigma_{Im_1}\sigma_{Im_2}} \tag{4.3}$$

where *cov* is the covariance normalized by σ being the standard deviation, ensuring $|corr| \leq 1$. Based on this similarity measure we further compared spin- and VISS-images in table 4.1 for three different sets of point-pairs in between the floating and target patch. The first set consists of corresponding points both of them not laying in the vicinity of a patch border. The second set is similar, but one of the two points within a pair lays close to a patch border. The last set consists of non-corresponding points. Each set contains ten point-pairs based on which the average correlation results in table 4.1 are determined. It is observed that VISS-images are capable of recognizing corresponding point-pairs even for points near borders. The latter are classified into non-corresponding points using the spin-images.

Given the multitude of possible correspondences between all point-signatures on the floating and target patches, a robust search procedure is required to filter an acceptable subset. Due to the partial overlap false correspondences are incorporated. It is thus important to distinguish correct from incorrect

	spin-images	VISS-images
corresponding point-pairs, no border point	0.8	0.95
corresponding point-pairs, one border point	0.5	0.9
non-corresponding point-pairs	0.45	0.7

Table 4.1: Averaged correlation coefficients for spin-images and VISS-images , between points of a floating and a target patch.

correspondences and to select the best correspondences, while determining a rigid transformation based on the correspondences, to recover an estimate of the relative position between the two patches. A simple and naive way is to select the best correspondences based on the similarity of the point-signatures. We use a pose clustering approach as proposed in [83]: for every possible rigid transformation that translates a point to its corresponding point and that makes their normal directions coincide, votes are accumulated. This is done for every established corresponding point-pair. Finally, the transformation that receives the maximum number of votes is selected. Using a voting scheme is robust, because every good corresponding point pair will contribute to the best transformation, while incorrect point pairs will vote on randomly different places in the voting scheme.

The reason for this registration method, based on directed VISS-images in combination with a voting scheme, being a crude registration or initialization is twofold. The most important reason is the quantization of the voting scheme such that the transformation is only correct within a quantization error. Secondly, to speed up computations, only a subset of all the points on the patches is used for finding correspondences based on the VISS-images. This makes the registration less accurate than fine registration methods. However, the method is robust and independent of the initial poses, due to the VISS-images combined with a voting scheme, which is applicable to a wide range of object surfaces besides faces.

4.2.2 Multi-view crude initialization

When more than two or multiple patches have to be initialized, a multi-view crude registration is needed. In this case the absolute pose estimate of every patch into a common coordinate frame, instead of pair-wise relative poses, is searched for, which is illustrated in figure 4.4. The common strategy for multi-view crude registration is to perform consecutive pair-wise crude registrations. However some new complications are introduced in a multi-view setup: (1) Which patch pairs will be registered? (2) Danger for creating inconsistent poses.

During pair-wise crude registration, relative pose estimates are calculated between every pair of patches. However, some registrations are better than others and some are even incorrect, because of the limited amount or lack of overlap. In multi-view crude registration, we select a non-redundant set of best

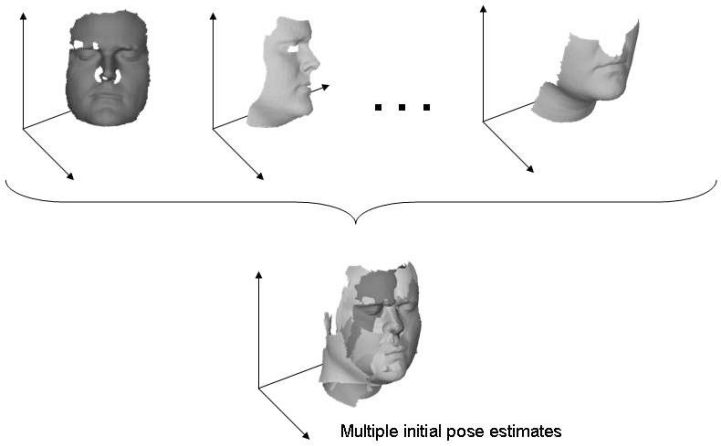


Figure 4.4: The purpose of multi-view crude registration: Given multiple patches within their own coordinate systems, bring all of them into a common coordinate system by applying initial pose estimates.

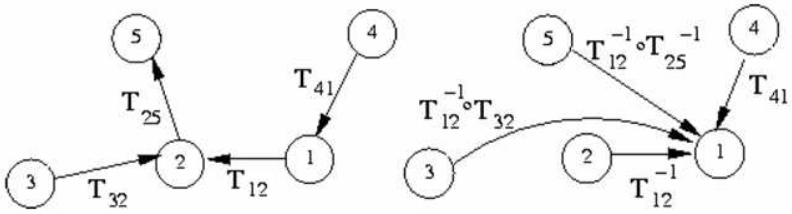


Figure 4.5: Illustration of the Minimum Spanning Tree algorithm

pair-wise registrations that gives an initial estimate of the absolute positions of all patches into the coordinate system of a reference patch. The quality of a single pair-wise registration is measured by the number of *transformation inliers*, an inlier being defined as a corresponding point-pair with an inter-point distance below a certain threshold. Inliers are corresponding point-pairs spatially supporting the pair-wise crude transformation and the number of inliers is a measure for or indication of the registration quality. More inliers indicate more acceptable corresponding point-pairs, due to the amount of overlap, resulting in a better transformation estimate. A minimum spanning tree algorithm (MST) is then used to select a (minimum) set of $P-1$ transformations to transform all P patches into a common coordinate system, maximizing the number of inliers. The MST has branches such that all patches are connected (figure 4.5) , while the sum of the costs is minimized, and does not generate cyclic connections that could create inconsistent poses.

4.2.3 Pair-wise fine registration

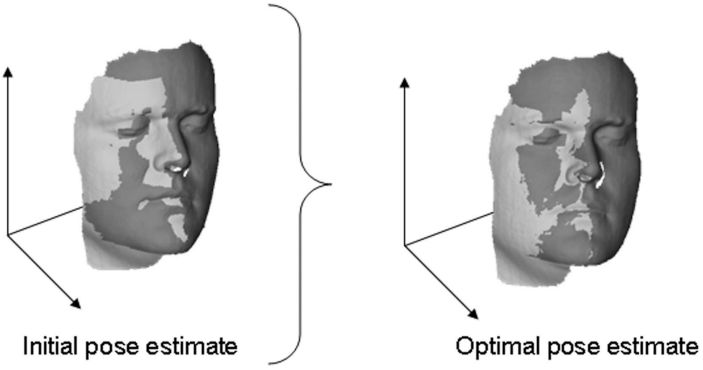


Figure 4.6: Given an initial pair-wise pose estimate between a floating and a target patch, the purpose of pair-wise fine registration is to refine the initial pose estimate into an optimal pose estimate.

To improve the initial estimates before the computationally expensive multi-view fine registration (section 4.2.4), we first apply computationally cheaper pair-wise fine registrations, illustrated in figure 4.6, on the minimum set of $P-1$ transformations obtained by the previous multi-view crude registration in the third step of the overall algorithm.

Given an initial pair-wise pose estimate between a floating and a target patch, the purpose of pair-wise fine registration is to refine the initial pose estimate into an optimal pose estimate. The most frequently used fine registration algorithm is the iterated closest point algorithm (ICP) introduced in [59] based on mesh representations of both the floating and target patches, where the distance between corresponding points, being closest points, is iteratively minimized. However, the basic ICP algorithm has several shortcomings. Firstly, when aligning point sets, both are typically two different samplings of the same shape. Even when the patches are perfectly aligned, the distance between a pair of best matching points can be nonzero. Secondly, the original algorithm in [59] was proposed for the registration of point sets with complete overlap. When dealing with partial overlap, many points of one patch will not have a corresponding partner on the other patch. However, points without partner do get matched to some point on the other patch. Such incorrect matches are referred to as outliers due to partial overlap. The final problem is the very expensive closest point search. A naive closest point search requires $N^f \times N^t$ point-to-point distance computations for every iteration, where N^f are the number of points on the floating patch and N^t are the number of points on the target patch.

In order to overcome the shortcomings of the basic ICP algorithm (cf. appendix D), we apply the robust surface registration framework from chapter 3.

The floating patch is represented as a set of points $S^f = \{\mathbf{r}_j^f | j = 1, \ldots, N^f\}$, while the target patch is represented by a variational implicit surface $S^t = \mathfrak{d}^t(\mathbf{r})$. Because no local surface deformations are necessary, a rigid transformation model, consisting of a translation and a rotation, suffice to register the floating patch with the target patch.

$$S^{f*} = \mathfrak{T}(S^f, \boldsymbol{\theta}) = \{\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) | j = 1, \ldots, N^f\} \tag{4.4}$$

with

$$\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) = R\mathbf{r}_j^f + \mathbf{t} \tag{4.5}$$

where R is a rotation matrix parameterized using euler angles ω_x, ω_y and ω_z constituting three consecutive rotations around the basis axes of the coordinate system:

$$R(\omega_x, \omega_y, \omega_z) = R_z(\omega_z) \circ R_y(\omega_y) \circ R_x(\omega_x) \tag{4.6}$$

and $\mathbf{t}$ is a translation vector according to the same basis axes:

$$\mathbf{t} = [t_x, t_y, t_z]^T \tag{4.7}$$

The complete set of rigid transformation parameters to estimate is $\boldsymbol{\theta} = [\omega_x, \omega_y, \omega_z, t_x, t_y, t_z]$, which is a six-dimensional transformation parameter vector. Due to the rigidity of the transformation the regularization $\|\mathfrak{L}(\boldsymbol{\theta})\|^2 = 0$, such that $p(\boldsymbol{\theta}) = cte$ in (3.15) for every possible transformation instance.

Using the rigid transformation model and the patch representations the individual point similarities in (3.18) are defined with predefined distances $d_j(\boldsymbol{\theta}) = 0$ independent of the parameters $\boldsymbol{\theta}$, ensuring the floating surface to be aligned onto the target surface:

$$\mathfrak{s}_j(\boldsymbol{\theta}) = \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) \quad \text{for} \quad j = 1, \ldots, N^f \tag{4.8}$$

Note that no explicit point correspondences are used in (4.8) in contrast to ICP and (3.17). The similarity measures and the transformation model are combined into an objective function according to the MAP formulation of section 3.3.3. Using (3.24), (3.25) and the fact that $\log p(\boldsymbol{\theta}) = cte$ we see that the noise σ estimation is irrelevant for optimization over $\boldsymbol{\theta}$ and can be omitted. Noise in the target surface is coped with using the smoothing properties of variational implicit functions during VIS fitting. Furthermore, the log model-likelihood term in (3.24) vanishes. The simplified objective function without outlier handling becomes:

$$\mathfrak{E}(\boldsymbol{\theta}) = \sum_{j=1}^{N^f} \mathfrak{s}_j(\boldsymbol{\theta})^2 \tag{4.9}$$

Proceeding with the theory from section 3.3.3, outliers are dealt with by introducing a latent variable outlier map $\mathbf{z}$ and an outlier-process. Assuming that outliers are solely due to the partial overlap, a regular latent variable with very strong prior knowledge as in (B.19) based on the variational boundary

function (VBF) $\mathfrak{b}^t(\mathbf{r})$ of the open target patch surface defined in section 3.2.3 can be applied. This assumption is valid for the facial patches we are working with. However, explained in section 3.3.3, a combined deterministic and probabilistic latent variable can be used in case the assumption is incorrect. Thanks to the interpolation and extrapolation properties of $\mathfrak{b}^t(\mathbf{r})$, points in 3D located directly above and below the patch surface S^t will evaluate positive in $\mathfrak{b}^t(\mathbf{r})$, while points next to and far away (outliers) will evaluate negative. Indeed, starting in a point in the middle of S^t and approaching the border or boundary of S^t, the function $\mathfrak{b}^t(\mathbf{r})$ will gradually decrease to zero. When traversing the border this decrease will continue and $\mathfrak{b}^t(\mathbf{r})$ becomes negative. Outliers due to partial overlap will evaluate negative, while inliers will evaluate positive in $\mathfrak{b}^t(\mathbf{r})$. Therefore, a deterministic outlier detection function (B.19) can be defined as:

$$z_j = f(\mathbf{r}_j^f(\boldsymbol{\theta})) = H\left(\mathfrak{b}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) - \mathrm{t}\right) \tag{4.10}$$

with $H(x')$ the Heaviside function ($H(x') = 0$ for $x' < 0$, $H(x') = 1$ for $x' > 0$). The threshold t is relatively easy to determine compared to other outlier detection and rejection thresholds found in literature (cf appendix D), since it only depends on the target patch and the constraints that create the function $\mathfrak{b}^t(\mathbf{r})$. A threshold t equal to or close to zero will be appropriate. Note that, using (4.10) leads to a binary variable $z_j = \{0, 1\}$: $\mathbf{r}_j^f$ is either in or out.

The resulting MAP objective function with outlier handling becomes:

$$\mathfrak{E}(\boldsymbol{\theta}, \mathbf{z}) = \sum_{j=1}^{N^f} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j)) \tag{4.11}$$

where

$$h(x, \alpha(z), \beta(z)) = \alpha(z)x^2 + \beta(z) \tag{4.12}$$

with

$$\alpha(z) = z \quad \beta(z) = -(1 - z)\log\delta \tag{4.13}$$

With δ the uniform outlier distribution parameter. The difference of (4.12) and (4.13) with (B.7) and (B.8) is the elimination of σ as an extra parameter, due to reasons given previously.

As mentioned in section 3.3.4, the optimal pose estimate solution can be found through an iterative joint optimization, leading to a dual updating procedure similar to an EM optimization. A sequence of parameter estimates $\{\widehat{\boldsymbol{\theta}}^{(m)} | m = 0, 1, \ldots\}$ is obtained by alternating the following two steps:

LV-step: On the $(m+1)^{th}$ iteration, the latent variables are eliminated by simple deterministic function evaluations using (4.10). Similar to (B.20) and (B.21) the following equivalent M-estimator is created:

$$\rho(x, \alpha, \beta) = h(x, \alpha(H(x')), \beta(H(x')) \tag{4.14}$$

such that

$$\mathfrak{E}(\boldsymbol{\theta}) = \sum_{j=1}^{N^f} \rho(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha, \beta) \tag{4.15}$$

with

$$\rho(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha, \beta) = h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^m))), \beta(f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^m)))) \tag{4.16}$$

M-step The ML parameters $\widehat{\boldsymbol{\theta}}^{(m+1)}$ are then updated minimizing the energy in (4.15) with fixed latent variable values or equivalently using (4.5), (4.8), (4.10), (4.12) and (4.13):

$$\mathfrak{E}(\boldsymbol{\theta}) = \sum_{j=1}^{N^f} \left[f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^m)) \left[\mathfrak{d}^t(R\mathbf{r}_j^f + \mathbf{t}) \right]^2 - (1 - f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^m))) \log \delta \right] \tag{4.17}$$

The ML parameters should satisfy $\partial\mathfrak{E}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$ and are updated similar to (3.44). Due to the smooth and analytical properties of VIS representations, first derivatives of (4.17) towards every parameter can always be computed and are determined analytically. The actual minimization of (4.17) is accomplished through a six dimensional conjugate gradient search framework instead of the regular Steepest-Descent gradient search in (3.44):

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - s\nabla_{conj}^{(m+1)}\mathfrak{E}(\boldsymbol{\theta}) \tag{4.18}$$

with $\nabla_{conj}^{(m+1)}\mathfrak{E}(\boldsymbol{\theta})$ the conjugate gradient of $\mathfrak{E}(\boldsymbol{\theta})$ on the $m+1^{th}$ iteration computed using regular gradient ∇ information, according to:

$$\nabla_{conj}^{(m+1)} = \nabla^{(m+1)} + \gamma^{(m)}\nabla_{conj}^{(m)} \tag{4.19}$$

with

$$\gamma^{(m)} = \frac{\left(\nabla^{(m+1)} - \nabla^{(m)}\right) \cdot \nabla^{(m+1)}}{\nabla^{(m)} \cdot \nabla^{(m)}} \tag{4.20}$$

Old gradient information is taken into account in order to avoid the zig-zagging behavior of regular Steepest-Descent algorithms and as a consequence faster convergence is obtained. The initial conjugate gradient at the first iteration $m = 0$ is set equal to the regular gradient:$\nabla_{conj}^{(1)} = \nabla^{(1)}$. The optimal step size s in the direction of the conjugate gradient is determined using a line minimization. The first derivatives for gradient calculations are straightforwardly computed using the Chain-Rule. Note that the term $(1 - f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^m))) \log \delta$ in (4.17) originating from the outlier-process distribution is constant during the M-step, therefore vanishing during derivative calculations. As a result, the choice of δ is of low (no) importance, when using a deterministic outlier detection function. However, having a combined random and regular latent variable z_j, the choice of δ is of importance for the random variable during the LV-step.

After applying pair-wise fine registration, the patches can be transformed into a common coordinate system based on the $P-1$ *refined* relative transformations and the MST. However, when dealing with noisy data and partial overlap, pair-wise registration errors will be accumulated in this way. To correct these errors we eventually apply a multi-view simultaneous fine registration as explained in the next section.

4.2.4 Multi-view fine registration

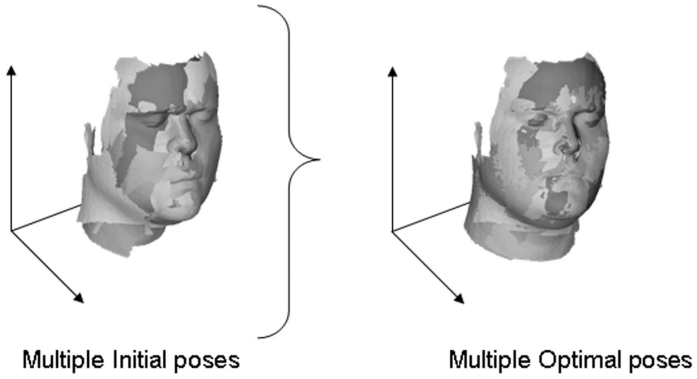


Figure 4.7: Illustration of multi-view fine registration. Given a set of initial absolute pose estimates , find the optimal absolute pose estimates

The purpose of multi-view fine registration is depicted in figure 4.7 and is the last registration step to accomplish. The common strategy for multi-view fine registration is similar to the crude case and is to perform consecutive pair-wise fine registrations. However, again some complications are introduced compared to the pair-wise fine registration: (1) building up of consecutive pair-wise registration errors and (2) inconsistent poses. The latter is solved using the MST from section 4.2.2. To cope with the former complication, a simultaneous registration of a patch with every other patch can be performed. Doing so, doesn't suffer from error-build up, but the complexity grows exponentially with the number of patches. Alternatively, an intermediate integration of all the simultaneous target patches can be used to register a floating patch like [81, 82].

Our multi-view fine registration is very similar to the pair-wise fine registration. The only difference between them is the target patch used for registration. In the pair-wise case the target patch consisted of one other patch, while in the multi-view case the target patch is an integrated version of multiple patches except for the floating patch itself. In fact integration and registration are merged into one optimization step like [81, 82], while traditionally they were performed separately. The integration of different patches $\{S_k^t | k = 1, \ldots, P-1\}$ into one

integrated target surface S^{int} with integrated VIS $\mathfrak{D}(\mathbf{r})$ representation is done with the integration method described in section 4.3.

The transformation model remains unchanged and is defined in (4.5). The individual point similarities defined in (4.8) are measured using an integrated signed distance surface $\mathfrak{D}(\mathbf{r})$ representation. When having a sufficient number of patches covering an object or when incorporating the floating patch to be registered into the integrated target patch, no outlier detection and rejection is needed during registration of a floating patch, because total overlap instead of partial overlap is present. However when not having enough patches or the floating patch is not incorporated into the model patch, outliers need to be accounted for using the individual patch weighting functions $\mathfrak{w}_k^*(\mathbf{r})$ for integration defined in (4.25) of section 4.3.1. Instead of (4.10), the latent variable is then determined using the following deterministic outlier detection function:

$$z_j = f(\mathbf{r}_j^f(\boldsymbol{\theta})) = H\left(\max\{\mathfrak{w}_k^*(\mathbf{r}_j^f)|k = 1, \ldots, P-1\}\right) \tag{4.21}$$

Replacing, (4.10) with (4.21) and $\mathfrak{d}^t(\mathbf{r})$ with $\mathfrak{D}(\mathbf{r})$, the LV-step (4.15) and M-step (4.17) are obtained. The same conjugate optimization strategy is used, by taking the derivative of (4.17) with respect to $\boldsymbol{\theta}$ and setting it equal to zero. The derivative of the function $\mathfrak{D}(\mathbf{r})$ in a point $\mathbf{r}$, needed for optimization, is more complicated, but can still be evaluated analytically (4.28).

To conclude, the complete algorithm for multi-view fine registration is described in pseudo-code:

Initialize
for $k = 1, \ldots, P$ **do**
  calculate $\mathfrak{d}_k^t(\mathbf{r})$ and $\mathfrak{w}_k^*(\mathbf{r})$ from S_k^t
end
procedure
repeat
  for $k = 1, \ldots, P$ **do**
    calculate $S^{int} = \mathfrak{D}(\mathbf{r})$ based on all but one patches (section 4.3.1)
    register S_k^t with S^{int} based on $\mathfrak{D}(\mathbf{r})$,(4.21), (4.15) and (4.17)
    transform S_k^t, $\mathfrak{d}_k^t(\mathbf{r})$ and $\mathfrak{w}_k^*(\mathbf{r})$
  end
until convergence;

It must be noted that the rigid transformation of a variational implicit function is straightforward. The basis of the polynomial term together with the location of the constraint points need to be transformed, while the RBF coefficients remain the same. No extra variational implicit function fitting is required.

4.3 Patch integration

Besides registration, all the patches are to be integrated into a single facial surface or entity. This is accomplished during the integration phase elaborated in this section, which is merged with the multi-view fine registration from section 4.2.4. Furthermore, besides shape integration an additional texture integration is required as well. Shape and texture integration within this section are based on weighted averages of variational implicit surfaces (VIS) and variational texture functions (VTF), respectively. The weighting functions, which give a local measure of confidence, are also constructed using the VIF machinery from section 3.2.1 based on confidence weight constraints, generating variational weighting functions (VWF). In section 4.3.1 the shape integration is elaborated, while section 4.3.2 focuses on the texture integration. For an overview of existing integration algorithms we refer the interested reader to appendix D.

4.3.1 Shape integration

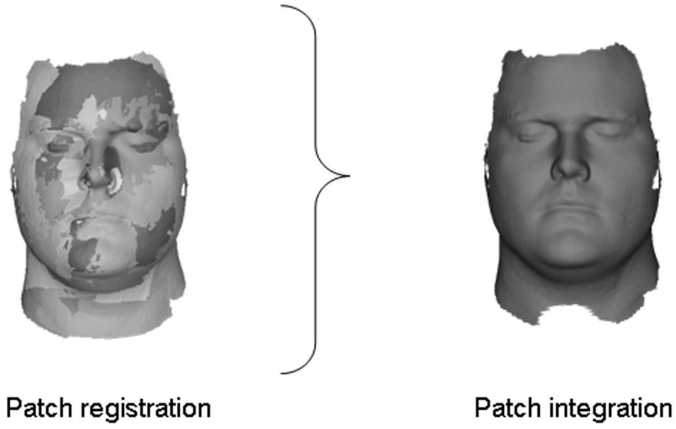


Figure 4.8: The purpose of shape integration is to combine multiple overlapping patches into a single complete surface, while taking into account local patch quality.

The purpose of shape integration is to combine multiple overlapping patch surfaces into a single complete surface, while taking into account local patch quality, shown in figure 4.8. Surface integration methods differ by the type of input data used, unorganized or connected point sets, and the type of surface representation, parametric or implicit. The most successful and popular integration techniques are the volumetric (implicit) integration methods as was first proposed by [85].

Volumetric integration of different patches $\{S_k | k = 1, \ldots, P\}$ into a single

surface, as proposed by [85], uses a discrete, grid-based, implicit function $D(\mathbf{r})$, representing a weighted average of signed distances to multiple patches. The resulting integrated surface is defined as the zero-valued iso-surface $D(\mathbf{r}) = 0$.

$$D(\mathbf{r}) = \frac{\sum_{k=1}^{P} w_k(\mathbf{r}) d_k(\mathbf{r})}{\sum_{k=1}^{P} w_k(\mathbf{r})} \tag{4.22}$$

$d_k(\mathbf{r})$ is a signed distance function of patch $S_k = \{\mathbf{r} | d_k(\mathbf{r}) = 0\}$. $w_k(\mathbf{r})$ is a weight function, which locally gives a measure of confidence for the function $d_k(\mathbf{r})$, where the confidence measures are derived from [86] for optical triangulation scanners. The purpose of the weight functions is to combine surface points measured within multiple patches while taking into account the individual quality of the multiple measurements, resulting in a weighted combination of measurements from a single surface point. The functions $d_k(\mathbf{r})$ and $w_k(\mathbf{r})$ are constructed on a regular three-dimensional grid, by shooting rays through every grid-point and the 3D sensor location of the patch acquisition. Final surface tessellation or mesh extraction from (4.22) is obtained using Marching Cubes [87].

Interesting properties of this volumetric integration (e.g. outlier robustness, incremental update, incorporation of range uncertainty, ...) are enlisted in [85]. However, several shortcomings and issues are to be dealt with. Firstly, a disadvantage of the ray shooting algorithm is the dependency of the resulting signed distance functions on the sensor location, which may not be known. It is also preferable to incorporate patch normals during build up of the distance functions in order to cope with sharp structures in the patch shape. Secondly, a major drawback of these traditional volumetric integration methods is the use of a 3D grid, such that the final surface resolution is limited by memory requirements and computational complexity. Thirdly, it still remains unanswered to what extent grid-based signed distance build-up is sensitive to the presence of noise within the patch data. Is the final signed distance value in each grid-point reliable, when the input patch is very noisy (larger than the grid sampling size,e.g.) and does this affect the final integration? Especially patch normal information is severely influenced by noise. Building up the signed distance function with incorporation of these corrupted normals will lead to a wrong integration result. Finally, every surface integration method also has to cope with the presence of holes in some of the patches. This is typically solved after integration and mesh extraction or during integration requiring additional overhead and/or grid-point classifications. No volumetric integration method uses a hole filling inherent representation.

We tackle the volumetric shape integration problem using the VIF machinery from sections 3.2.1, retaining the desirable properties enlisted in [85], but at the same time resolving current limitations of volumetric integration. In a first instance, the signed distance functions $d_k(\mathbf{r})$ in (4.22) for every patch are replaced with the VIS representations of the patches $\mathfrak{d}_k(\mathbf{r})$. In a second instance, the weight functions $w_k(\mathbf{r})$ are replaced by variational implicit weight functions (VWF) $\mathfrak{w}_k^*(\mathbf{r})$ based on range uncertainty or point confidences de-

fined in [85]. The border ∂S_k of every patch S_k and the distance $d(\mathbf{r}_{ki}, \partial S_k)$ of every point $\mathbf{r}_{ki}$ in S_k to the border for $i = 1, \ldots, K_k$, with K the number of points on patch S_k, are determined. For every 3D point $\mathbf{r}_{ki}$ of a patch S_k a corresponding weight w_{ki} (point confidence) is calculated according to:

$$\begin{aligned} c'_{ki} &= w' \log(1 + w'' d(\mathbf{r}_{ki}, \partial S_k)) \\ c''_{ki} &= \tfrac{2}{\pi} \arcsin(\mathbf{n}_{ki}, \mathbf{s}_k) \\ w_{ki} &= c'_{ki} c''_{ki} \end{aligned} \tag{4.23}$$

c'_{ki} gives lower weights to points near a patch border, because borders of patches are often less reliable. Note that these weights are exactly the same as the boundary constraints in section 3.2.3, with (w', w'') equal to $(1/\log(100), 99/(\max_i d(\mathbf{r}_{ki}, \partial S_k)))$. c'_{ki} takes into account the angle between the normal $\mathbf{n}_{ki}$ in a point of a patch S_k and the sensor direction $\mathbf{s}_k$ from which the patch was captured. Points at grazing angles receive lower weights. If the sensor direction $\mathbf{s}_k$ is not known, it is approximated by the average normal of all the points on the patch S_k:

$$\mathbf{s}_k = \frac{1}{K_k} \sum_{i=1}^{K_k} \frac{\mathbf{n}_{ki}}{\|\mathbf{n}_{ki}\|} \tag{4.24}$$

w_{ki} combines c'_{ki} and c''_{ki}. Figure 4.9 shows c'_{ki} (right), c''_{ki} (middle) and the combined weights w_{ki} (left) on an exemplar patch.

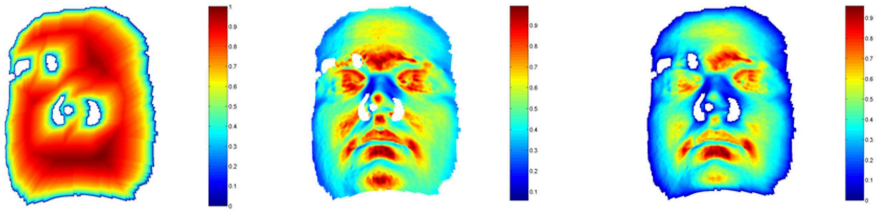


Figure 4.9: c' (right), c'' (middle) and the combined weights w (left) from (4.23) on an exemplar patch. The color-maps range from 0 (blue) to 1 (red).

Based on the points $\{\mathbf{r}_{ki} | i = 1, \ldots, K_k\}$ of a patch S_k, being the constraint points, and there weights $\{w_{ki} | i = 1, \ldots, K_k\}$, being the function constraints, a new variational implicit function $\mathfrak{w}_k(\mathbf{r})$ can be constructed (section 3.2.1). Doing so, local on-patch confidence information is extrapolated to off-patch positions (see figure 4.11) and combined with the distance function $\mathfrak{d}_k(\mathbf{r})$ a new weight function $\mathfrak{w}^*_k(\mathbf{r})$ per patch can be defined:

$$\mathfrak{w}^*_k(\mathbf{r}) = \mathfrak{w}_k(\mathbf{r}) H\left(\mathrm{t_d} - |\mathfrak{d}_k(\mathbf{r})|\right) H\left(\mathfrak{w}_k(\mathbf{r}) - \mathrm{t_w}\right) \tag{4.25}$$

with $H(x)$ the Heaviside function. The threshold $\mathrm{t_d}$ is introduced to eliminate the influence of a patch to points in front or behind the patch which are

further away than t_d, so that thin objects can be reconstructed. Two patches at opposite sides of an object should not influence each other. The threshold t_w is the same as t for the outlier handling of section 4.2.3 in (4.10) and limits the influence of a patch outside his border. Hole filling consists in a good determination of this threshold t_w or by not taking into account borders of small holes in the patches for the determination of the distances $d(\mathbf{r}_{ki}, \partial S_k)$ in (4.23). Using $\mathfrak{w}_k^*(\mathbf{r})$ and $\mathfrak{d}_k(\mathbf{r})$ instead of $w_k(\mathbf{r})$ and $d_k(\mathbf{r})$ in (4.22) we obtain:

$$\mathfrak{D}(\mathbf{r}) = \frac{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})\mathfrak{d}_k(\mathbf{r})}{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})} \tag{4.26}$$

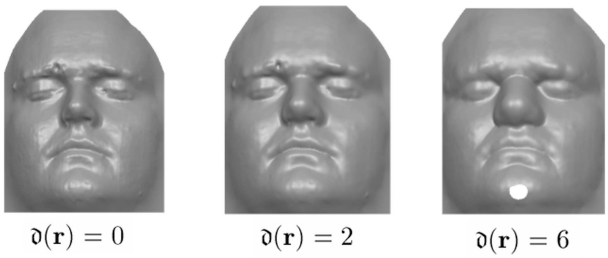


Figure 4.10: Three mesh tessellations of the exemplar patch VIS $\mathfrak{d}(\mathbf{r})$, (left) mesh for $\mathfrak{d}(\mathbf{r}) = 0$, which is the original surface with extrapolations and interpolations. (middle) mesh at distance $\mathfrak{d}(\mathbf{r}) = 2$. (right) mesh at distance $\mathfrak{d}(\mathbf{r}) = 6$. These tessellations are used to visualize the extrapolation properties of the VWF and VTF in figure 4.11 and 4.13, respectively.

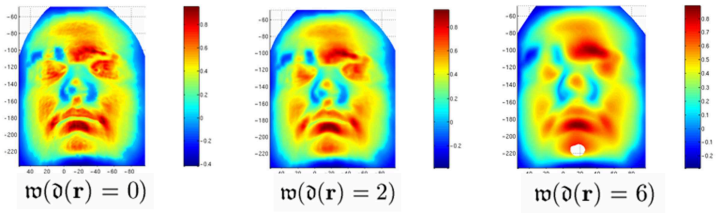


Figure 4.11: Visualization of the VWF extrapolation properties, based on the tessellations in figure 4.10. Confidence weight information on the patch surface is smoothly extrapolated towards points off the surface, which is needed for integration.

A final mesh tessellation of the zero iso-distance surface $\mathfrak{D}(\mathbf{r}) = 0$ is obtained using Marching Triangles (MT) [88]. The most interesting property of MT compared to Marching Cubes (MC) is that no 3D grid is needed. Only, three requirements are needed to use MT. Firstly, a seed triangle is needed to initialize the growing. One or more triangles of the original patches can be

used and projected on the zero iso-surface of $\mathfrak{D}(\mathbf{r})$ as seed triangles. Secondly, evaluation of the implicit function must be possible for every point in 3D. This is done with (4.26). Finally, it must be possible to determine the derivative of the implicit function in a 3D point. To calculate the derivative of $\mathfrak{D}(\mathbf{r})$ in an arbitrary 3D point $\mathbf{r}$ the patches that have an influence on the point are to be determined.

$$\mathfrak{K} = \{k | H\left(\mathrm{t_d} - |\mathfrak{d}_k(\mathbf{r})|\right) H\left(\mathfrak{w}_k(\mathbf{r}) - \mathrm{t_w}\right) = 1, \text{ for } k = 1, \ldots, P\} \tag{4.27}$$

Then, the derivative of $\mathfrak{D}(\mathbf{x})$ based on the subset patches $\mathfrak{K}$ is calculated analytically as:

$$\frac{\partial \mathfrak{D}(\mathbf{r})}{\partial \mathbf{r}} = \frac{A_1 - A_2}{B^2} \tag{4.28}$$

with

$$\begin{aligned} A_1 &= \textstyle\sum_{k \in \mathfrak{K}} \frac{\partial \mathfrak{w}_k(\mathbf{r})}{\partial \mathbf{r}} \mathfrak{d}_k(\mathbf{r}) + \sum_{k \in \mathfrak{K}} \mathfrak{w}_k(\mathbf{r}) \frac{\partial \mathfrak{d}_k(\mathbf{r})}{\partial \mathbf{r}} \\ A_2 &= \textstyle\sum_{k \in \mathfrak{K}} \frac{\partial \mathfrak{w}_k(\mathbf{r})}{\partial \mathbf{r}} \sum_{k \in \mathfrak{K}} \mathfrak{w}_k(\mathbf{r}) \mathfrak{d}_k(\mathbf{r}) \\ B &= \textstyle\sum_{k \in \mathfrak{K}} \mathfrak{w}_k(\mathbf{r}) \end{aligned} \tag{4.29}$$

By making use of variational implicit surfaces and weight functions of the patch surfaces and weights some limitations of current volumetric integration methods are overcome. No memory inefficient 3D grid is used, such that the final resolution or accuracy is not limited by memory requirements. Hole filling does not require an extra step because of the interpolation and extrapolation properties of the representations. Furthermore, the variational representations are sensor location independent and take into account surface normal information during representation build-up. Finally, useful smoothing properties of VIF [69] can be used when dealing with noisy patch data.

4.3.2 Texture integration

In a final stage of the complete face acquisition algorithm from partial surfaces, a texturing of the integrated surface obtained in section 4.3.1 (after applying MT) is performed using texture integration, which is depicted in figure 4.12. The amount of literature concerning texture integration or blending is limited compared to shape integration. Texture contributions from different patches are typically weighted during blending. Disadvantages of current texture integrations methods [89, 90, 91] are the need for integrated surface extraction before texture weight determination and the dependency of the weights on the relative poses of the patches, which makes the texture weights and blending sensitive to the accuracy of the patch registrations.

Our texture integration is performed in the same way as the shape integration of the previous section. Instead of a weighted combination of VIS representations, three weighted combinations of variational implicit texture functions (VTF) is performed. The weight functions are the same as for shape

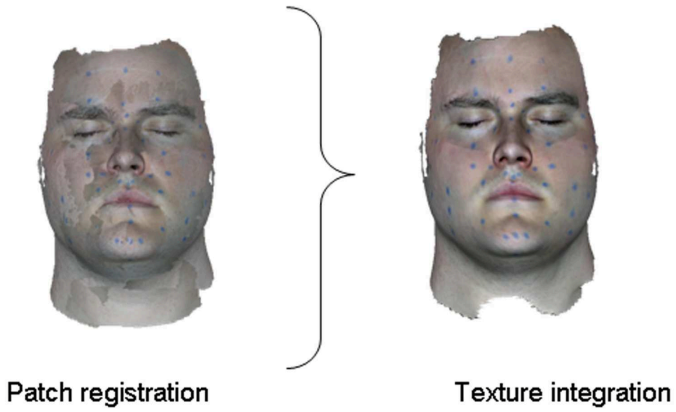


Figure 4.12: The purpose of texture integration is to combine multiple overlapping textures into a single smooth texture map.

integration, being independent on the relative poses and therefore being independent of registration errors. Furthermore, no integrated surface extraction is needed prior to texture weight computation.

During patch acquisition, for every point $\{\mathbf{r}_{ki}|i = 1, \ldots, K_k\}$ of every patch $\{S_k|k = 1, \ldots, P\}$ an RGB triplet (r_{ki}, g_{ki}, b_{ki}) is acquired besides 3D location information. Three VTFs $\mathfrak{t}_k^r(\mathbf{r})$, $\mathfrak{t}_k^g(\mathbf{r})$ and $\mathfrak{t}_k^b(\mathbf{r})$ for every patch S^k are constructed based on the constraint points $\{\mathbf{r}_{ki}|i = 1, \ldots, K_k\}$ and function constraints

$$\begin{aligned} \{\mathfrak{t}_k^r(\mathbf{r}_{ki}) &= r_{ki}|i = 1, \ldots, K_k\} \\ \{\mathfrak{t}_k^g(\mathbf{r}_{ki}) &= g_{ki}|i = 1, \ldots, K_k\} \\ \{\mathfrak{t}_k^b(\mathbf{r}_{ki}) &= b_{ki}|i = 1, \ldots, K_k\} \end{aligned} \tag{4.30}$$

using the VIF machinery from section 3.2.1. Doing so, on-patch texture information is extrapolated towards off-patch locations (see figure 4.13). Using $\mathfrak{t}_k^r(\mathbf{r})$, $\mathfrak{t}_k^g(\mathbf{r})$ and $\mathfrak{t}_k^b(\mathbf{r})$, integrated texture functions $\mathfrak{R}(\mathbf{r})$, $\mathfrak{G}(\mathbf{r})$ and $\mathfrak{B}(\mathbf{r})$ are obtained based on weighted sums similar to the shape integration in (4.26):

$$\begin{aligned} \mathfrak{R}(\mathbf{r}) &= \frac{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})\mathfrak{t}_k^r(\mathbf{r})}{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})} \\ \mathfrak{G}(\mathbf{r}) &= \frac{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})\mathfrak{t}_k^g(\mathbf{r})}{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})} \\ \mathfrak{B}(\mathbf{r}) &= \frac{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})\mathfrak{t}_k^b(\mathbf{r})}{\sum_{k=1}^{P} \mathfrak{w}_k^*(\mathbf{r})} \end{aligned} \tag{4.31}$$

Texturing of the integrated surface mesh obtained using MT on (4.26), is then accomplished by evaluating every mesh vertex of the integrated surface

into $\mathfrak{R}(\mathbf{r})$, $\mathfrak{G}(\mathbf{r})$ and $\mathfrak{B}(\mathbf{r})$ of (4.31), resulting in an integrated RGB triplet for every mesh vertex. By making use of variational implicit texture and weight functions, the same advantages enumerated in section 4.3 are valid for the texture integration as well.

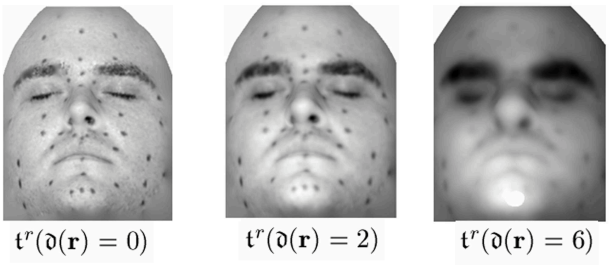


Figure 4.13: Visualization of the VTF extrapolation properties, based on the tessellations in figure 4.10. Texture information on the patch surface is smoothly extrapolated towards points off the surface, which is needed for integration. In this example the red VTF $\mathfrak{t}^r(\mathbf{r})$ is used.

4.4 Results

This section shows some experimental results of our 4-step registration and integration algorithm on real-live, noiseless and noisy simulated data. The data was acquired using a 3D active range scanning camera (ShapeCam, Eyetronics, Leuven, Belgium [92]) and associated ShapeWare reconstruction software. Human faces are fairly complicated objects to acquire. The nose typically causes a lot of problems, because of the amount of curvature and being partially occluded in different patches. Hair and eyebrows introduce severe noise and local surface misinterpretations. A last difficulty is the non-rigidity, like swallowing, of the face during acquisition of different patches, which introduces imperfect surface matches between patches.

Two different aspects are analyzed: robustness in terms of initialization (being completely automatic) and accuracy. Robustness in terms of initialization, is mainly determined by the crude registration, while accuracy is mainly dependent on the fine registration. The crude registration is successful if it outputs a good initialization for the fine registration to converge to an accurate result. The feasibility of the algorithm is shown in section 4.4.1 on real live-data. Based on this data set, a further comparison between VISS-images and spin-images is made. Furthermore, the integration power is illustrated as well, based on very erroneous individual real-live patch acquisitions. Due to the lack of ground truth knowledge of the individual patches within the real-live data set we create a simulated data set in section 4.4.2, to analyze accuracy in terms of registration and transformation parameter errors. Using the simulated data set, the pair-wise fine registration of section 4.2.4 is also

briefly compared with an ICP fine registration technique [93]. Furthermore, the same data set is applied to show a texture integration of artificial higher frequency individual patch textures compared to regular non dynamic skin texture. To illustrate a strategy based on VIFs to handle noise, a final noisy data set is created in section 4.4.3.

4.4.1 Real-live data set

The first data set used is a real-live face acquisition, shown in figures 4.14 and 4.15, consisting of 11 patches. Notice the very bad nose reconstruction in the close-up images of figure 4.14, together with holes typically located at the nose and the eyebrows. Observe the wrongfully aligned textures in some patches of figure 4.15, due to a time-delay between texture and geometry image (e.g. figure 4.1(middle)) capturing for a single patch acquisition. Normally, the individual patches are first cleaned up manually with the ShapeWare software, which involves smoothing, hole filling, texture alignment and deletion of bad parts. We choose not to perform the manual cleanup to reduce the manual work to a minimum and to illustrate the feasibility and robustness of our method to bad input data.

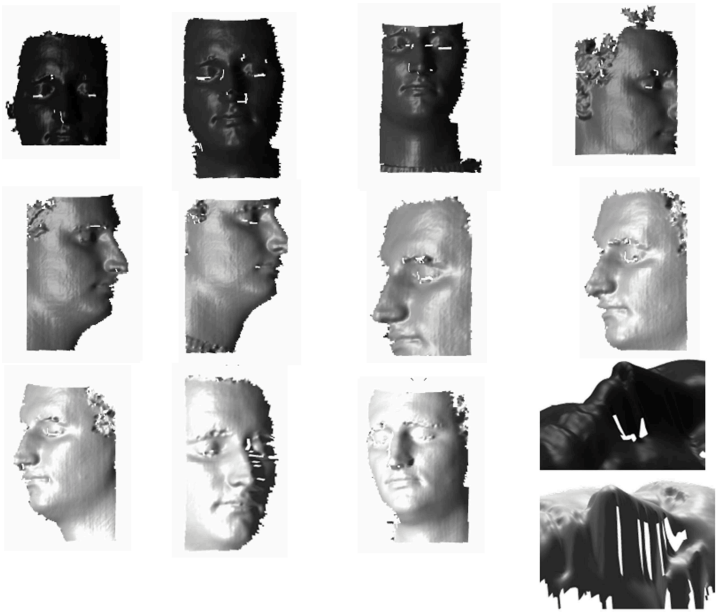


Figure 4.14: Real-live data set consisting of 11 patches. Included are two close-ups of the nose in two different patches to illustrate bad nose acquisitions due to misinterpretations and occlusions during acquisition.

For the pair-wise crude registration, we selected randomly 500 points on every patch for which VISS-images were computed. The VISS-image resolution

Figure 4.15: 11 patches of the real-live data set rendered with texture information. Note the bad texture alignments in some patches, due to a time-delay between texture and geometry image (e.g. figure 4.1(middle))acquisitions.

was 15×15 bins with a bin size of 2 by 2 mm. Other VISS-image parameters are: $s = 36$, $\gamma = 4$. For the voting scheme, three quantization resolutions are defined. The translation vector has quantization intervals of $20mm$, the rotation vector has intervals of 0.2 mm and the rotation angle has intervals of 0.0066 rad. This results in a very coarse six-dimensional voting scheme when converting the rotation vector and angle into euler angles. Despite the low resolution of the VISS-images and low quantization resolution of the voting scheme, the crude registration manages to initialize the fine registrations well enough in order for them to converge successfully as can be seen in figure 4.16. Of course by increasing both resolutions better crude registrations can be obtained at the expense of computation time and memory.

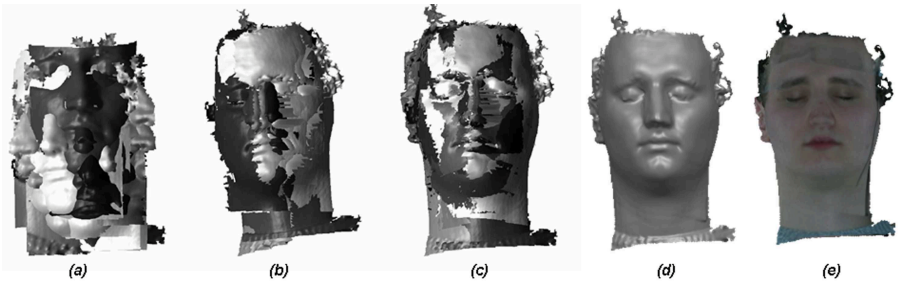


Figure 4.16: Progress of the algorithm on the real-live data set: (a) initial situation, (b) after multi-view crude registration, (c) after pair-wise and multi-view fine registration, (d) surface integration , (e) texture integration.

Transf.	VISS	Spin
$\mathfrak{T}_1$	213	139
$\mathfrak{T}_2$	176	132
$\mathfrak{T}_3$	144	101
$\mathfrak{T}_4$	138	96
$\mathfrak{T}_5$	128	77
$\mathfrak{T}_6$	121	70
$\mathfrak{T}_7$	117	70
$\mathfrak{T}_8$	117	66
$\mathfrak{T}_9$	105	59
$\mathfrak{T}_{10}$	104	58
Avg(1-10)	**136.3**	**86.8**
Avg(1-110)	**61.5**	**37.5**

Table 4.2: The number of inliers per transformation ($\mathfrak{T}$), found by VISS-images and spin-images

In section 4.2.1 some advantages of the VISS-images compared to spin-images were listed and shortly illustrated. Here, a further comparison is made based on the number of transformation supporting inliers (cf. section 4.2.2) obtained with both point-signatures using the same resolution parameters. Being an inlier or correct point-pair is defined as: An established corresponding point-pair obtained through normalized correlation between point-signatures, lying in each others proximity after applying the computed crude transformation. Proximity of point-pairs is limited to $15mm$ according to the translation resolution of $20mm$. Doing so, an indication on the number of correct point-pairs is obtained and given in table 4.2. The average amount of inliers over the $P-1=10$ with $P=11$ best transformations selected after multi-view crude registration and over all possible 110 pair-wise transformations are given as well in table 4.2. It is clearly observed that comparing VISS-images performes better than spin-images in terms of the amount of transformation supporting inliers. The average over all possible transformations is much lower for both spin- and VISS-images, but one has to keep in mind that not all patch-pairs share a common overlap. Additionally, from the 110 transformations, we visually examined the amount of satisfying transformations found by the VISS-images (72 out of 110) and spin-images (53 out of 110). 72 for the VISS-images is considered to be good knowing that not all 110 transformations are valid or realistic.

A visual result of the integration power is given in figure 4.17. A close-up of the final integrated nose is given. Keeping in mind the bad partial nose acquisitions in figure 4.14, we see that the result is quite acceptable, but not yet perfect. This can be resolved by deleting some really bad nose parts in the individual patches, like in the closeup images of figure 4.14. Alternatively, more patches could be acquired so that more measurements of the same data

is incorporated, downgrading the contribution of erroneous measurements. A third option is to define a weight function similar to c'_{ki} and c''_{ki} in (4.23), that detects bad parts like the close-ups in figure 4.14 and gives them a lower weight during integration. Due to the lack of ground truth knowledge no further analysis can be performed. The quantification of accuracy is obtained in the next section, using a simulated data set with ground truth incorporation.

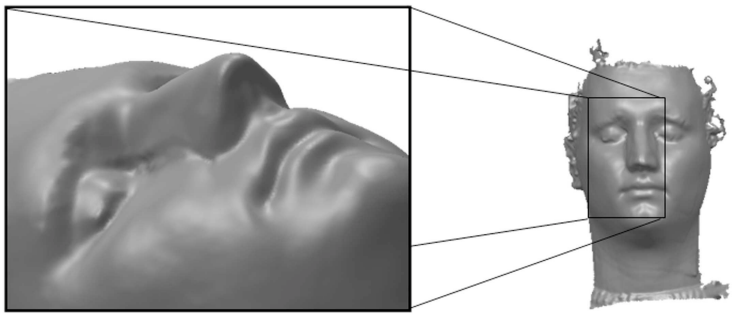


Figure 4.17: Close-up of the nose after integration

4.4.2 Simulated data set

The second data set is a noiseless simulated data set illustrated in figure 4.18. Starting from a complete 3D surface of a face obtained using the ShapeWare software and some additional manual editing, 10 partially overlapping patches were cut out of the 3D surface. Subsequently, all patches except one were randomly translated from 0 to 50 mm in every direction and rotated between 0 and 10 degrees around every axis in 3D. The goal of the registration process is to align the patches into the coordinate system of the unchanged patch. By doing so we can compare the registration per patch with the ground truth, being the starting model which lies in the same coordinate system. The registration accuracy per patch can then be calculated as the mean and the standard deviation in mm of the absolute function evaluations of the patch points into the variational implicit surface of the original complete 3D surface. The crude registration parameters are the same as for the previous data set. The threshold t and t_w needed during pair-wise and multi-view fine registration respectively, are set to 0. The influence threshold t_d is set to 15 mm.

Figure 4.19 depicts the progress of the algorithm on the noiseless simulated data set. The absolute local surface differences or registration errors, obtained using absolute VIS evaluations, between the final integrated result in figure 4.19(d) and the ground truth surface in figure 4.18(a) are shown in figure 4.18(e) using a gray-map ranging from 0 mm (black) to 1 mm (white). It is observed that the final result is more than satisfying in terms of local surface accuracies. The average and standard deviation of the local surface differences

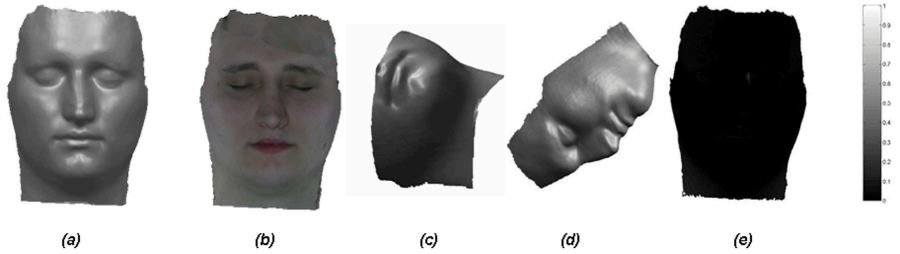


Figure 4.18: Simulated data set: (a) Ground truth shape. (b) Ground truth texture. (c) and (d) two exemplar patches of 10 patches obtained from the complete surface. (e) Local comparison of the final integration in figure 4.19 (d) with the ground truth surface (a) using absolute VIS evaluations

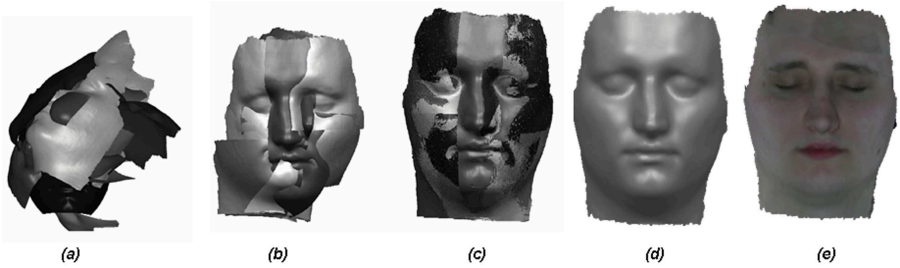


Figure 4.19: Progress of the algorithm on the simulated data set: (a) initial situation, (b) after multi-view crude registration, (c) after pair-wise and multi-view fine registration, (d) surface integration , (e) texture integration.

after integration are 0.0223 mm and 0.0361 mm respectively. Similarly, individual patch averages and standard deviations of local registration errors after the different algorithmic steps are enumerated in table 4.3. Thanks to the notion of pair-wise ground truth transformation parameters, remaining transformation parameter errors after registration can be computed as in table 4.4 for pair-wise fine registrations after the $P-1=9$ extracted crude registrations using the MST method during the multi-view crude registration step.

In table 4.3, it is observed that the mean absolute error over all patches (except the first) drops from 8.9 mm (multi-view crude) over 0.03 mm (pair-wise fine) to 0.01 mm (multi-view fine). The first patch S_1 was not altered from its ground truth position. Therefore small registration errors are obtained for the first patch, which are not incorporated into the total registration error calculations. Taking into account table 4.4, while analyzing table 4.3, we see that the pair-wise fine registration severely increases the registration quality. The multi-view fine registration further reduces the registration error, but the reduction is not severe because of the already nice results obtained after the pair-wise fine registrations. It is therefore interesting to apply pair-wise

Patch	MV Crude		PW fine		MV fine	
	μ	σ	μ	σ	μ	σ
S_1	3.6e-6	2.9e-6	3.6e-6	2.9e-6	3.6e-6	2.9e-6
S_2	16.844	7.1940	0.0383	0.0277	0.0185	0.0173
S_3	8.0178	4.8434	0.0505	0.0405	0.0185	0.0179
S_4	7.6327	3.0579	0.0630	0.0655	0.0200	0.0152
S_5	3.0873	2.0597	0.0015	0.0013	0.0034	0.0020
S_6	11.902	5.4532	0.0355	0.0316	0.0175	0.0175
S_7	6.0803	3.7855	0.0014	0.0012	0.0035	0.0022
S_8	7.6714	2.8056	0.0013	0.0010	0.0043	0.0028
S_9	7.2438	4.1538	0.0313	0.0300	0.0240	0.0224
S_{10}	10.8515	5.4650	0.0353	0.0269	0.0131	0.0141
$Total$	**8.8990**	**5.9924**	**0.0287**	**0.0381**	**0.0137**	**0.0161**

Table 4.3: Averages μ and standard deviations σ in mm of the local surface registration errors for every patch $\{S_k|k = 1, \ldots, 10\}$ separately and all the patches (except the first patch S_1) combined (Total).

	$\mathfrak{T}_1$	$\mathfrak{T}_2$	$\mathfrak{T}_3$	$\mathfrak{T}_4$	$\mathfrak{T}_5$	$\mathfrak{T}_6$	$\mathfrak{T}_7$	$\mathfrak{T}_8$	$\mathfrak{T}_9$
t_x	-0.004	-0.095	0.052	0.001	-0.001	0.003	<1e-3	0.089	-0.007
t_y	-0.013	0.099	-0.220	-0.002	0.008	-0.002	0.003	-0.075	0.009
t_z	-0.011	0.210	0.187	0.005	0.004	-0.005	0.005	0.040	0.002
ω_x	<1e-3	-0.001	0.002	<1e-3	<1e-3	<1e-3	<1e-3	0.001	<1e-3
ω_y	<1e-3	0.002	-0.001	<1e-3	<1e-3	<1e-3	<1e-3	0.001	-<1e-3
ω_z	<1e-3	<1e-3	<1e-3	<1e-3	<1e-3	<1e-3	<1e-3	0.002	<1e-3

Table 4.4: Remaining pair-wise transformation parameter $\boldsymbol{\theta} = [t_x, t_y, t_z, \omega_x, \omega_y, \omega_z]$ errors for pair-wise fine transformation estimates $\{\mathfrak{T}_k|k = 1, \ldots, 9\}$ after the $P - 1 = 9$ extracted crude registrations using the MST method during the multi-view crude registration step. t_x, t_y and t_z in mm. ω_x, ω_y and ω_z in *rad* (radial).

fine registrations after multi-view crude and before multi-view fine registration, because the latter is computationally more expensive. The multi-view fine registration only needed 5 iterations before convergence. Comparing the pair-wise fine registration error and the multi-view fine registration error intra patch-wise, it is seen that the error is mostly reduced but is sometimes slightly increased. This copes with the main idea of multi-view simultaneous registration trying not only to reduce but also to redistribute pair-wise registration errors.

Starting from the $P - 1$ selected best pair-wise crude registrations using the MST during multi-view registration, we initialize both our pair-wise fine registration algorithm from section 4.2.3 and an ICP related fine registration [93]. The ICP algorithm uses point-based surface representations for both the target and the floating surface and outliers are dealt with using a maximum distance threshold in between corresponding points. One of the pair-wise cases is shown in figure 4.20. The starting position after pair-wise crude registration is shown in 4.20(b). The ICP and our result are shown in 4.20(a) and (c) respectively. For both results local surface differences are visualized using a

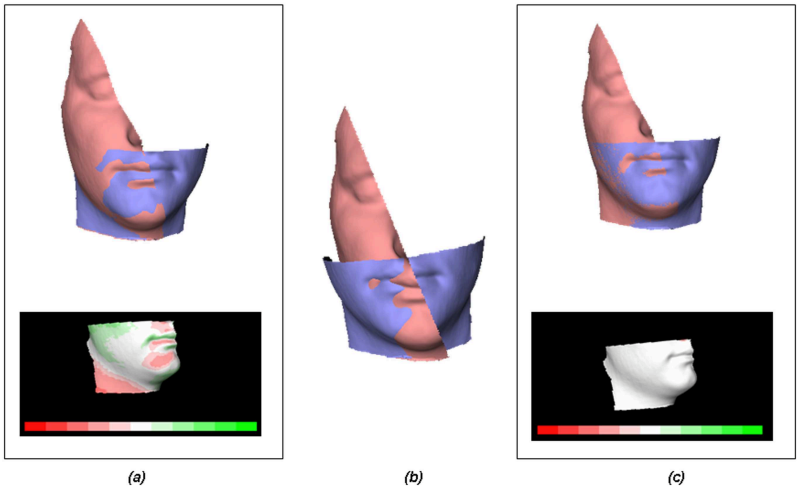


Figure 4.20: (c) The pair-wise fine registration of section 4.2.3 based on VIS target surface representations and a deterministic outlier detection function, compared with an ICP related fine registration (a), both starting from the same crude initialization (b)

color-code ranging from -3 mm (red) to 3 mm (green). The outlier threshold for the ICP algorithm was determined using a trial and error approach after which the best result was retained. It is seen that the registration accuracy with our registration algorithm is superior. The other 8 cases generated similar results. However, it must be noted that the used ICP algorithm is not one of the best to deal with partial overlap, but it was the only implementation available at the time of validation. Furthermore, a huge variety of ICP algorithms exist in literature, making it difficult to choose a golden standard ICP algorithm to compare with.

A final illustration (see figure 4.21) based on the simulated data set, from this section, concerns the ability of the texture integration, elaborated in section 4.3.2, to cope with high frequency or highly detailed textures. Typical face textures are of low variation. Therefore, the facial texture information from the ground truth facial surface in figure 4.18 is replaced with artificial texture incorporating higher color variance and the textures of the individual patches are altered accordingly as well. The shape integration result from figure 4.19(d) is evaluated within the integrated texture functions (4.31) to obtain the textured result in figure 4.21(a). Comparing the result with the ground truth 4.21(b), we can conclude that the texture integration is indeed capable to cope with high variance containing textures, besides regular facial textures.

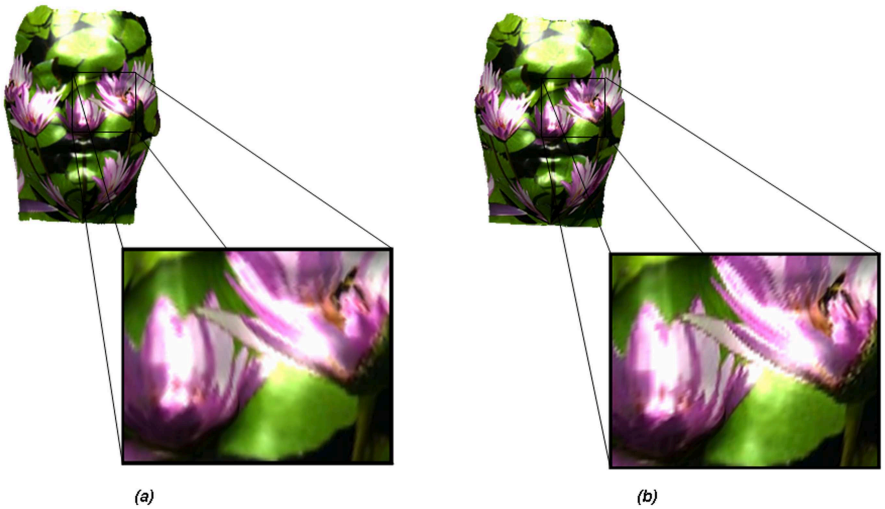


Figure 4.21: Illustration of the texture integration ability to cope with high dynamic textures. (a) reconstructed texture after texture integration with close-up. (b)ground truth texture with close-up.

4.4.3 Noisy simulated data set

The third and final data set is a noisy synthetic data set (see figure 4.22) constructed in the same way as the previous data set, except that we add noise to the patch point coordinates. The noise is normally (gaussian) distributed with zero mean and 0.5 mm standard deviation. The purpose of this data set is to show that the smoothing properties of the variational implicit surfaces render the registration scheme robust to noise. The smoothing is introduced into all 4 registration steps by substituting a smoothed VIS $\mathfrak{d}_k^s(\mathbf{r})$ for the original VIS $\mathfrak{d}_k(\mathbf{r})$. To create the functions $\mathfrak{d}_k^s(\mathbf{r})$ from the noisy data we choose to use the spline smoothing technique during fitting described in section 3.2.1 using (3.11), with a smoothing parameter $\nu = 32$. Figure 4.23 depicts the progress of the algorithm on the noisy simulated data set.

VISS-images are mainly determined by the variational implicit surface $\mathfrak{d}_k(\mathbf{r})$ of the patch S_k, the location of the point $\mathbf{r}$ and its normal direction from which the plane I is constructed. Especially the function $\mathfrak{d}_k(\mathbf{r})$ and the normal directions have a major influence on the images, more than the point locations. The latter only has a small translation effect, while the former two have a more complex and a rotation effect, respectively. Creating a smooth function $\mathfrak{d}_k^s(\mathbf{r})$ reduces the noise influence on the VIS representation. To correct the normal directions from noise influences the gradient of the smoothed $\mathfrak{d}_k^s(\mathbf{r})$ is computed in every patch point and used as smoothed normal. Finally, the noisy patch vertex or point locations can be corrected by projecting the points locally onto the zero-distance surface of $\mathfrak{d}_k^s(\mathbf{r})$ using (3.46). However,

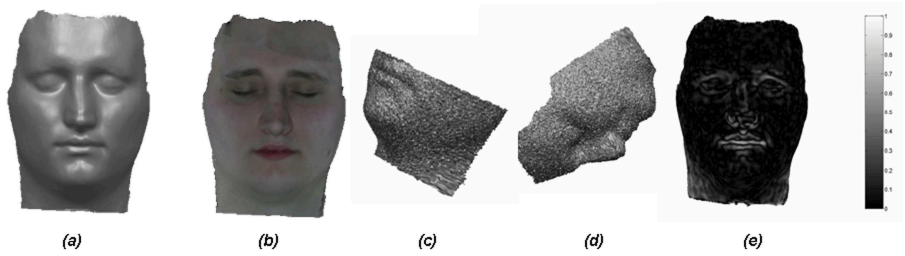


Figure 4.22: Noisy Simulated data set: (a) Ground truth shape. (b) Noiseless Ground truth texture. (c) and (d) two noisy exemplar patches of 10 patches obtained from the complete surface and additive gaussian noise. (e) Local comparison of the final integration in figure 4.23 (d) with the noiseless ground truth surface in (a) using absolute VIS evaluations

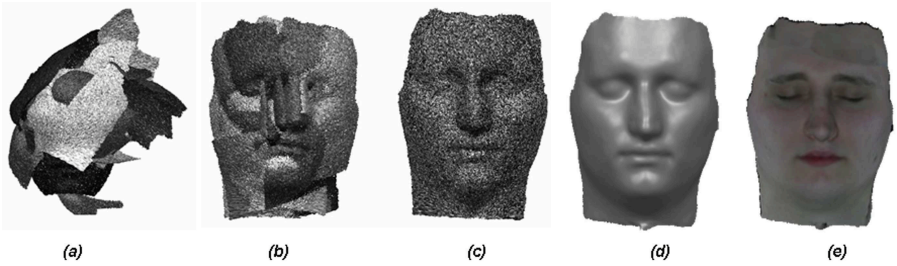


Figure 4.23: Progress of the algorithm on the noisy simulated data set: (a) initial situation, (b) after multi-view crude registration, (c) after pair-wise and multi-view fine registration, (d) surface integration , (e) texture integration.

this latter correction concerning the point locations is not done, because crude registration succeeded, by experiment, without it. Using spin-images instead of the smoothed VISS-images failed to give proper crude initializations, due to the sensitive nature of spin-images towards noisy mesh representations. A spin-image is dependent on the exact location of a point $\mathbf{r}$ with its normal and neighboring points, which are all corrupted by noise. In order for spin-images to succeed a noiseless mesh representation is to be extracted from the noisy mesh representation.

Using $\mathfrak{d}_k^s(\mathbf{r})$ instead of $\mathfrak{d}_k(\mathbf{r})$ in (4.8) and (4.26) render the pair-wise and multi-view fine registrations, respectively, robust against noise. Doing so, gradient computations, needed for the conjugate-gradient optimization, are now noiseless and smooth. This is comparable with the signed Distance Transform determination based on filtered normals in [81]. Incorporation of a noise estimation in (4.9) would not resolve the problem of noisy gradient calculations. Furthermore, such a noise estimation only influences the amount of transformation regularization, which is always zero for rigid transformations. Therefore the noise estimation in (4.9) was not needed and was neglected. Fi-

Patch	MV Crude		PW fine		MV fine		Noise	
	μ	σ	μ	σ	μ	σ	μ	σ
S_1	0.413	0.312	0.413	0.312	0.413	0.312	0.414	0.313
S_2	8.629	3.563	0.4391	0.3313	0.4163	0.3156	0.4096	0.31
S_3	7.582	4.628	0.426	0.321	0.425	0.321	0.411	0.311
S_4	14.55	3.970	0.418	0.317	0.413	0.312	0.411	0.310
S_5	3.878	2.420	0.411	0.312	0.411	0.312	0.410	0.312
S_6	6.478	3.307	0.413	0.312	0.416	0.314	0.409	0.309
S_7	10.94	5.763	0.4179	0.313	0.416	0.312	0.414	0.310
S_8	10.94	6.145	0.413	0.329	0.416	0.318	0.409	0.314
S_9	4.541	2.867	0.413	0.313	0.415	0.315	0.409	0.309
S_{10}	4.541	2.867	0.413	0.313	0.415	0.315	0.409	0.309
$Total$	**8.353**	**5.308**	**0.420**	**0.318**	**0.415**	**0.314**	**0.410**	**0.310**

Table 4.5: Averages μ and standard deviations σ in mm of the local surface registration errors for every noisy patch $\{S_k|k = 1, \ldots, 10\}$ separately and all the patches (except the first patch S_1) combined (Total).

nally, substituting $\mathfrak{d}_k(\mathbf{r})$ with the smooth function $\mathfrak{d}_k^s(\mathbf{r})$ in (4.26) also results in a smooth noise free final surface shape integration.

The absolute local surface differences between the final integrated result in figure 4.23(d) and the noiseless ground truth surface in figure 4.22(a) are shown in figure 4.22(e) using a gray-map ranging from 0 mm (black) to 1 mm (white). It is observed that the final result is smooth instead of noisy and is more than satisfying in terms of local surface accuracies. The average and standard deviation of the local surface differences after integration are 0.0864 mm and 0.0901 mm respectively. Similar to the previous section using the noiseless synthetic data set, registration error and transformation error analyzes per patch are made in table 4.5 and table 4.6 respectively. The mean absolute error (by comparing the registrations of the noisy patches to a noiseless ground truth) over all patches (except the first) drops from 8.4 mm (multi-view crude) over 0.420 mm (pair-wise fine) to 0.416 mm (multi-view fine). The remaining registration error is mainly equal and due to the noise, with average absolute noise level of 0.41 mm, so that we can conclude that the registration was performed robustly and successfully. Comparing these numbers with the accuracy after integration we see that a major amount of the noise present in the patch data is eliminated during integration. Finally, the transformation parameter errors in table 4.6 after pair-wise fine registrations are satisfying and acceptable in practise.

Note that, the simulated noise we added does not necessarily occur in reality, but when knowing the noise distribution of the 3D acquisition system or the noise per patch data point, one of the smoothing techniques, besides spline-smoothing, enlisted in [69] can be applied. As a consequence, the main strategy and robustness of the algorithm against noise, does not change depending on the type of noise.

	$\mathfrak{T}_1$	$\mathfrak{T}_2$	$\mathfrak{T}_3$	$\mathfrak{T}_4$	$\mathfrak{T}_5$	$\mathfrak{T}_6$	$\mathfrak{T}_7$	$\mathfrak{T}_8$	$\mathfrak{T}_9$
t_x	-0.097	0.274	0.143	0.073	-0.021	0.235	-0.006	0.064	-0.057
t_y	0.007	0.455	-0.382	-0.014	0.083	-0.196	-0.058	0.049	0.019
t_z	-0.088	0.392	0.179	0.078	-0.031	-0.022	0.035	0.003	-0.014
ω_x	<1e-3	<1e-3	0.003	-0.001	-0.002	0.005	-0.001	<1e-3	-0.001
ω_y	-0.001	0.007	-0.001	-0.001	0.001	0.002	<1e-3	<1e-3	-0.002
ω_z	0.001	-0.001	<1e-3	<1e-3	-0.001	0.007	<1e-3	-0.004	<1e-3

Table 4.6: Remaining pair-wise transformation parameter $\boldsymbol{\theta} = [t_x, t_y, t_z, \omega_x, \omega_y, \omega_z]$ errors for pair-wise fine transformation estimates $\{\mathfrak{T}_k | k = 1, \ldots, 9\}$ on the noisy patches after the $P - 1 = 9$ extracted crude registrations using the MST method during the multi-view crude registration step. t_x, t_y and t_z in mm. ω_x, ω_y and ω_z in rad (radial).

4.5 Conclusion

Facial surfaces as part of the craniofacial database were acquired using a photogrammetric active 3D acquisition device. Advantages included, portability and upright position acquisition. However, due to the limited viewing-angle of the camera a complete facial surface needed to be assembled from multiple partial surface acquisitions. To build a complete database an automatic assembling procedure for the facial surface acquisition was required.

We presented an automatic, robust and accurate registration and integration algorithm for 3D surface building from separate patches, without any prior knowledge. We used a variational implicit surface representation of the patches, which is a volumetric representation not requiring a memory inefficient and resolution-limiting 3D grid. The VIS representation had useful properties for solving the different tasks and had interesting smoothing properties to deal with noisy data. First, a pair-wise crude registration was solved by making use of directed point-signatures in combination with a voting scheme. Shortcomings of spin-images for use as point-signatures were solved by creating VISS-images increasing the quality and the robustness of the point-signatures. A minimum spanning tree (MST) was used to convert the pair-wise crude registrations into a multi-view crude registration resulting in a set of best transformations, according to the number of transformation inliers, needed to transform all the patches into a common coordinate system. This set of best transformations was further refined by applying pair-wise fine registrations. We defined an objective function, according to chapter 3, based on the VIS representation of the target patch. The optimization was accomplished by a six dimensional conjugate gradient search for the transformation parameters, instead of finding corresponding points between floating patch and target patch. We also defined a novel variational implicit function for an improved outlier detection and rejection during optimization. In contrast to traditional outlier detection systems, the outlier detection function depended on the target patch only, resulting in an easier outlier threshold setting. Additionally, a combined random/deterministic outlier variable could be used when outliers

are not solely introduced by partial overlap. Finally, accumulation of pair-wise registration errors was dealt with by applying a multi-view simultaneous fine registration. Solving this registration was similar to the pair-wise fine registration except that the target patch was replaced by an intermediate volumetric integration of patches. The integration itself was also based on the VIS representations such that integration and registration could be unified.

By making use of variational implicit surface and function representations of the patch surfaces, textures and weights we solved limitations of current volumetric integration methods. No memory inefficient 3D grid was used, such that the final resolution or accuracy is not limited by memory requirements. Hole filling did not require an extra step because of the interpolation and extrapolation properties of the chosen representations. Furthermore, the representation had useful smoothing properties when dealing with noisy input data. Texture integration was performed similarly to shape integration, making use of the same weight functions in order to reduce the texture influence of low confidence patch points.

We performed experiments on real-live data, noiseless and noisy simulated data. The real-live data set showed the robustness of the algorithm to bad data. In order to improve the final quality of the 3D model, some really bad data parts in the patches were to be deleted or more patches are to be taken in order to reduce the effect of the bad parts in the final integration. The noiseless data set showed the accuracy of the algorithm. The noisy simulated data set illustrated that the smoothing properties of the VIS representation can be used to make the algorithm robust against noisy input data.

Having acquired the complete facial surface and a set of ultrasound thickness measurements, the final prerequisite for a facial entry within the craniofacial database is to establish dense point correspondences with a reference face. The complete facial surface registration is dealt with in the next chapter.

Chapter 5

Facial surface registration

5.1 Introduction

Why inter-subject facial surface registration? The third and last prerequisite for our craniofacial database is to register facial surfaces to perform a statistical inter-subject facial analysis (chapter 6) and to construct a statistical craniofacial model. As a result of the patch integration and the final tessellation presented in the previous chapter, every facial surface in the database is represented as a dense set of connected 3D points. However, the acquired faces within the database, until now, do not have the same amount of 3D points sharing the same connectivity and corresponding texture information. To tackle this problem, the 3D points with known connectivity of a carefully constructed generic reference face are mapped or non-rigidly registered onto the faces in the database, which was already illustrated in figure 1.3. Using the definitions of chapter 3, the generic reference face is the floating surface while the faces in the database are the target surfaces. The result of inter-subject facial surface registration is that for every point on one surface the corresponding point on every other surface in the database is known. Furthermore, the 52 face landmarks, in which ultrasound tissue depths were measured, are a subset of the dense points, such that corresponding inter-subject facial anatomical landmarks are known as well. Gathering texture information within the registered points is accomplished by evaluating the points within the integrated variational texture functions (4.31) of the target faces.

Requirements The facial surface registration methodology must be robust against noise and must be able to cope with outliers due to partial overlap or missing data. Missing data are due to the fact that the 3D surface acquisition device is not able to acquire 3D information from hairy or covered regions in the face, generating incomplete 3D surface acquisitions with holes or missing data. An example is visible in figure 1.3: due to the presence of sidewhiskers, the 3D facial surface contains holes or misses 3D surface information

at the location of the whiskers. Noise robustness is not necessarily required for small errors in the target surface, because these were smoothed after shape integration in section 4.3.1, but rather because of the non-rigid nature of the registration. Indeed, non-rigid surface registration requires an iterative procedure alternating between finding corresponding points and calculation of the transformation geometrically aligning corresponding points. Hence, an intermediate estimate of the point correspondences during registration is typically erroneous, especially at the beginning of the registration, where the shapes of the floating and the target surface are still very different.

Shortcomings of existing methods The establishment of dense point correspondences using a non-rigid surface registration in between two surfaces is fairly complicated due to the the non-rigid nature of the registration problem. Furthermore, it is a well known fact that the quality of the statistical model extracted afterwards is dependent on the quality and the consistency of the established dense point correspondences. A short overview of related work concerning non-rigid facial surface registration for shape modeling is given in appendix F. Several authors simplify the non-rigid registration problem by user intervention to obtain a better initialization [94, 95, 96]. However, the final result is then dependent on the quality of the user intervention guiding the registration. Furthermore, it is time-consuming when applied to a large database. Alternatively, [97, 98] solve the problem in 2D instead of 3D using a 3D-2D mapping. However, a 3D-2D mapping incorporates the danger of loosing 3D surface information, especially in regions with high curvature. Therefore, we prefer to solve the problem in 3D instead. A non-rigid registration technique in 3D without user intervention is the robust point matching (RPM) algorithm [71], which we mention here, because it is closely related to our non-rigid surface registration method. The RPM algorithm is designed to establish correspondences between two point-sets and to determine a non-rigid Thin Plate Sline (TPS) based transformation mapping the corresponding point sets onto each other. Basically, two techniques are applied to tackle the non-rigid registration, soft-assign (for correspondence updating) and deterministic annealing (for TPS transformation updating). Thanks to the latter, severe shape differences between the initial floating point set and the target point set are slowly but deterministically reduced, such that no user-intervention is needed while remaining in 3D. In theory, the RPM algorithm could be directly applied to the surface matching problem by representing both floating and target surfaces as dense point-sets. However, the original RPM cannot handle a huge amount of points which is shown in [99], due to the size of the soft-assign matrix which is dependent on the amount of floating and target points and due to a direct implementation of the TPS deformation. Furthermore, a point based target surface representation leads to sampling errors within the final correspondences similar to the ICP algorithm for rigid patch pose refinement.

Our approach Our surface registration method also uses a TPS based non-rigid transformation which is optimized similar to [71] and is embedded within the robust surface registration framework of chapter 3 to deal with partial overlap. A resemblance between the TPS deformation and the VIF theory of section 3.2.1 can be shown, such that the VIF fitting and evaluating approximation machinery can be applied to calculate and evaluate the TPS transformations, making the algorithm feasible for a vast amount of floating surface points. Furthermore, the target surface is represented using a VIS, such that no sampling errors are contained within the final correspondences and such that the establishment of corresponding points during registration for a large amount of points remains computationally feasible. The theoretical aspects of the non-rigid facial surface registration method are given in section 5.2, based on the robust registration framework using VIS representations from chapter 3. Section 5.3, briefly touches on the problem of selecting of reference face. Some results are depicted in section 5.4, while a possible validation is discussed.

5.2 Facial surface registration framework

To tackle the problem of transforming a complete reference facial surface, being the floating surface, towards a possible incomplete facial surface within the database, being the target surface, a proper non-rigid TPS transformation model and similarity measure, based on VIS target surface representations, are chosen in section 5.2.1. Both the non-rigid transformation model and similarity measure are combined into a single objective function in section 5.2.2 using the framework of section 3.3.3 restating the registration problem into a MAP estimation problem. Outliers, due to partial overlap or missing data, are dealt with using a deterministic outlier detection function and probabilistic outlier-process, similar to the outlier handling during fine registrations of patches in section 4.2.3. Because of the non-convex nature of the object function a deterministic annealing optimization is applied, through deterministic manipulation of the noise estimate parameter.

Throughout this section the floating surface, being the reference face, is represented as set of points $S^f = \{\mathbf{r}_j^f | j = 1 \ldots, N^f\}$ and the target surface, being a facial surface from the database, is represented using a variational implicit surface (VIS) $S^t = \mathfrak{d}^t(\mathbf{r})$. Because faces in the database are open surfaces the VIS is augmented with a variational boundary function (VBF) $\mathfrak{b}^t(\mathbf{r})$. The goal of the facial surface registration is to find the corresponding geometrical relationship and to transform the floating surface towards the target surface. The latter can then be represented using a set of points $S^t = \{\mathbf{r}_j^t | j = 1 \ldots, N^f\}$ describing the target surface geometry, such that point couples $\{(\mathbf{r}_j^f, \mathbf{r}_j^t) | j = 1 \ldots, N^f\}$ are in correspondence and such that the amount of points on the target surface equals the amount of points on the floating surface while sharing the same connectivity.

5.2.1 Similarity measure and Transformation model

The similarity criterium should express that the floating surface point positions are optimal, when the points are on the zero-distance surface of the target VIS $\mathfrak{d}^t(\mathbf{r})$, such that they describe the target surface geometry. Therefore an individual point similarity measure as in section 4.2.3 with predefined distances $d_j(\boldsymbol{\theta}) = 0$ independent of $\boldsymbol{\theta}$ will suffice, ensuring the floating surface to be aligned onto the target surface:

$$\mathfrak{s}_j(\boldsymbol{\theta}) = \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) \quad \text{for} \quad j = 1, \ldots, N^f \tag{5.1}$$

A rigid transformation model $\mathfrak{T}(\mathbf{r}, \boldsymbol{\theta})$ as in section 4.2.3 will not suffice to tackle the facial surface registration problem. Floating surface points are then restricted to rotate and translate and therefore are not capable of describing the geometry or shape of the target facial surface of different subjects. Instead a non-rigid transformation model is needed, allowing local surface deformations, such that the geometry of the floating surface can be altered towards the shape of the target surface. We choose a generic non-rigid TPS transformation model obtained by a parametrization of the transformation as a weighted sum of functionals, being generic basis functions, similar to the variational implicit functions in (3.3) incorporating a mathematical regularization. The regularization includes a function smoothness measure, ensuring local surface consistencies after transformation and a smooth transition of the floating surface towards the target surface. The non-rigid TPS transformation model is mathematically well defined and is applicable to a wide range of surfaces from different objects due to their generic nature.

Using a homogenous coordinate notation $[r_x, r_y, r_z, 1]^{\mathrm{T}}$, with $^{\mathrm{T}}$ the vector transpose, for a 3D point $\mathbf{r}$ the TPS transformation according to [71] is defined as:

$$\mathbf{r}_j^{f*} = \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) = C\mathbf{r}_j^f + \boldsymbol{\phi}(\mathbf{r}_j^f)\Lambda \tag{5.2}$$

,where C is a 4x4 rigid transformation matrix in which rotation and translation are combined compared to (4.5). Additionally a scaling and skew can be incorporated as well, rendering the rigid transformation into an affine transformation. Λ is a $N^f \times 4$ non-rigid warping coefficient matrix. The $1 \times N^f$ vector $\phi(\mathbf{r}_j^f)$ is related to the thin-plate spline $N^f \times N^f$ kernel matrix Φ with component values $\phi_i(\mathbf{r}_j^f) = \|\mathbf{r}_j^f - \mathbf{r}_i^f\|$ for $i = 1, \ldots, N^f$. Using the spline notation in (5.2) the deformation is decomposed into a rigid (affine) and a non-rigid (non-affine) subspace resulting into the following transformation parameters $\boldsymbol{\theta} = [\{C_{ij} | i = 1, \ldots, 3; j = 1, \ldots, 4\}, \{\Lambda_{ij} | i = 1, \ldots, N^f; j = 1, \ldots, 3\}]$. The last row of $C = [0001]$ and the last column of Λ consist of fixed values equal to zero, due to the homogenous coordinate notations.

Equivalently, the TPS transformation can be decomposed into three separate transformations according to three coordinate axes $b \in \{x, y, z\}$:

$$\mathbf{r}_{bj}^{f*} = \mathfrak{m}_b(\mathbf{r}_j^f) = \mathbf{c}_b\mathbf{r}_j^f + \sum_{i=1}^{N^f} \lambda_{ib}\phi_i(\mathbf{r}_j^f) \text{ for } b \in \{x, y, z\} \tag{5.3}$$

with $\mathbf{c}_b$ a 1×4 vector, which is one of the first three rows of C in (5.2). Similarly , λ_{ib} is an element of one of the first three columns in Λ. Analyzing (5.3) a great resemblance with the definition of a variational implicit function in (3.3) is observed. Indeed, defining the polynomial term in (3.3) as $v(\mathbf{r}) = \mathbf{cr}$ and using a 3D TPS or biharmonic spline as RBF $\phi(\|\mathbf{r}-\mathbf{r}_i\|) = \|\mathbf{r}-\mathbf{r}_i\|$, (3.3) is exactly the same as (5.3). This suggest that the VIF machinery from section 3.2.1 can be used to determine non-rigid TPS transformation parameters based on constraint points and function constraints. Given a set of corresponding points $\{(\mathbf{r}_j^{f*}, \mathbf{r}_j^f)|i = 1, \ldots, N^f\}$, the constraint points are the floating surface points $\{\mathbf{r}_j^f|i = 1, \ldots, N^f\}$ and the function constraints are defined as $\{h_{bj} = r_{bj}^{f*}|i = 1, \ldots, N^f\}$ for $b \in \{x, y, z\}$. Three variational mapping functions (VMF), $\mathfrak{m}_b(\mathbf{r})$ with $b \in \{x, y, z\}$ can then be constructed according to (3.5). Using TPSs, a smooth transition of the function $\mathfrak{m}_b$ in between consecutive constraint points is obtained, while interpolating through the function constraints. Furthermore, the transformation regularization $\|\mathfrak{L}(\mathfrak{m})\|^2$ is the same as in (3.10) and when dealing with noisy or erroneous correspondences additional spline-smoothing according to (3.11) and (3.12) can be applied.

The transformation regularization using the notation of (5.2) is equivalent to:

$$\|\mathfrak{L}(\boldsymbol{\theta})\|^2 = \text{trace}(\Lambda^{\mathrm{T}}\Phi\Lambda) \text{ with } \Phi_{ij} = \|\mathbf{r}_i^f - \mathbf{r}_j^f\| \tag{5.4}$$

Note that the original floating surface points are used as constraint points in the TPS kernel Φ. So loosely speaking, the TPS kernel contains information about the floating point-set's internal structural relationships. In other words, the transformation model likelihood in (3.15) using (5.4), incorporates prior knowledge concerning the floating surface. Regularization comes down to constraining the non-rigid transformation subspace in (5.2), not to deform the internal structure of the floating surface too arbitrarily. Determination of the transformation parameters based on corresponding points $S^f = \{\mathbf{r}_j^f|j = 1 \ldots, N^f\}$ and $S^{f*} = \{\mathbf{r}_j^{f*}|j = 1 \ldots, N^f\}$ with additional spline-smoothing $\nu > 0$ using the notation of (5.2) is obtained solving the following closed form linear system similar to (3.12):

$$\begin{pmatrix} \Phi + \nu I & S^f \\ S^{f^{\mathrm{T}}} & 0 \end{pmatrix} \begin{pmatrix} \Lambda \\ C \end{pmatrix} = \begin{pmatrix} S^{f*} \\ 0 \end{pmatrix} \tag{5.5}$$

where S^f and S^{f*} are notated as $N^f \times 4$ matrices consisting of concatenated homogeneous points $\mathbf{r}_j^f$ and $\mathbf{r}_j^{f*}$ for $j = 1 \ldots, N^f$ in every row.

Due to the link between (5.2), (5.4) and (5.5) with the VIF machinery, properties of VIFs are also valid for the non-rigid TPS transformation model. Furthermore, the algorithmic improvements implemented in the FastRBF^{TM} library [68] to solve (5.5) and to evaluate (5.2) can be applied, making the registration method applicable for many floating surface points.

5.2.2 Objective function and Optimization

The non-rigid surface registration objective function, combining both the similarity measure and the non-rigid transformation model from the previous section, is obtained using the theoretical framework of section 3.3.3, while handling outliers due to missing data or partial overlap. Assuming that outliers are solely due to the partial overlap, a regular latent variable with very strong prior knowledge as in (B.19), which is exactly the same as (4.10), based on the variational boundary function (VBF) of the open target facial surface $\mathfrak{b}^t(\mathbf{r})$, defined in section 3.2.3, can be applied. This assumption is valid for the facial surfaces after facial acquisition. However, as explained in section 3.3.3, a combined deterministic and probabilistic latent variable can be used in case the assumption is incorrect. Defining the set of ML-parameters $\Theta = \{\boldsymbol{\theta}, \sigma\} = \{C, \Lambda, \sigma\}$, including the unknown noise standard deviation, the objective function becomes:

$$\mathfrak{E}(\Theta, \mathbf{z}) = \mathfrak{S}(\Theta, \mathbf{z}) + \nu \|\mathfrak{L}(\theta)\|^2 \tag{5.6}$$

The second term is the negative log model-likelihood and is defined in (5.4) for the non-rigid TPS based transformation model. The first term is the negative log complete data-likelihood, which is dependent on a regular outlier latent variable map $\mathbf{z}$ and is defined as:

$$\mathfrak{S}(\Theta, \mathbf{z}) = \sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j)) \tag{5.7}$$

where

$$h(x, \alpha(z), \beta(z)) = \alpha(z)x^2 + \beta(z) \tag{5.8}$$

with

$$\alpha(z) = \tfrac{1}{2\sigma^2} z \quad \beta(z) = -(1-z)\log\delta + z\log\sqrt{2\pi}\sigma \tag{5.9}$$

The optimal solution is found through a joint parameter estimation, leading to a dual updating procedure. A sequence of parameter estimates $\{\Theta^{(m)} | m = 0, 1, \ldots\}$ is obtained by alternating the following two steps:

LV-step: On the $(m+1)^{th}$ iteration, the latent variables are eliminated from the complete negative log data-likelihood $\mathfrak{S}(\Theta, \mathbf{z})$, by simple deterministic function evaluations using (4.10). Doing so, an equivalent M-estimator ρ is constructed such that

$$\mathfrak{S}(\Theta) = \sum_{j=1}^{N^f} \rho(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha, \beta) \tag{5.10}$$

with

$$\rho(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha, \beta) = h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^{(m)}))), \beta(f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^{(m)})))) \tag{5.11}$$

M-step The ML parameters $\widehat{\Theta}^{(m+1)}$ are then updated minimizing the energy in (5.6), where (5.7) is replaced by (5.10) with fixed latent variable values or equivalently using (5.2), (5.1), (4.10), (5.8) and (5.9):

$$\mathfrak{E}(\Theta) = \mathfrak{S}(\Theta) + \nu \text{ trace}(\Lambda^{\mathrm{T}}\Phi\Lambda) \tag{5.12}$$

with

$$\begin{aligned}\mathfrak{S}(\Theta) = \quad & \textstyle\sum_{j=1}^{N^f} f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^{(m)})) \left[\frac{1}{2\sigma^2}\left(\mathfrak{d}^t(C\mathbf{r}_j^f + \boldsymbol{\phi}(\mathbf{r}_j^f)\Lambda)\right)^2 + \log\sqrt{2\pi}\sigma\right] - \dots \\ & \textstyle\sum_{j=1}^{N^f}(1 - f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^{(m)})))\log\delta\end{aligned} \tag{5.13}$$

Optimization of (5.12) using the analytical Steepest-Descent approach requires the explicit construction of the TPS kernel Φ during line minimization after gradient derivation. However, the size of the spline kernel ($N^f \times N^f$) is dependent on the amount of floating surface points and rapidly becomes impractical to work with. Instead an ICP alike optimization is performed using the implicit to explicit conversion in (3.46). Additional motivations for doing this, are the vast amount of parameters to estimate and the availability of a closed form transformation parameter calculation (5.5). Furthermore, during transformation determination or fitting and evaluation the domain decomposition and multi-pole algorithmic improvements of the Fastrbf^{TM} toolbox can be applied, such that the spline kernel is never explicitly constructed keeping the computational complexity equal to $O(N^f \log N^f)$ instead of $O(N^{f^3})$. Making the optimization explicit instead of implicit comes down to replacing the similarity measure defined in (5.1) with a similarity measure based on point-correspondences similar to (3.17) with predefined distances $d_j = 0$:

$$\mathfrak{s}_j(\boldsymbol{\theta}) = \|\mathbf{r}_j^t - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\| \tag{5.14}$$

with $\mathbf{r}_j^t$ the corresponding point on the target surface S^t of the point $\mathbf{r}_j^f$ on the floating surface S^f. Doing so, the M-step itself becomes a dual update alternating between a correspondence update and a transformation and noise parameter update. First 5.12 is simplified by only retaining the inlier points $\mathbf{r}_j^f \in \mathfrak{J} = \{\mathbf{r}_j^f | z_j = f(\mathbf{r}_j^f(\widehat{\boldsymbol{\theta}}^m)) = 1 \wedge j = 1, \dots, N^f\}$ after the LV-step, due to the binary nature of the latent variable. Furthermore, the $\log\delta$ contribution of outliers is a constant after the LV-step, independent of the transformation parameters, therefore vanishing during the M-step optimization:

$$\mathfrak{E}(\Theta) = \sum_{j\in\mathfrak{J}}\left[\frac{1}{2\sigma^2}\|\mathbf{r}_j^t - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\|^2 + \log\sqrt{2\pi}\sigma\right] + \nu \text{ trace}(\Lambda^{\mathrm{T}}\Phi\Lambda) \tag{5.15}$$

To solve this energy function, first a set of correspondences is to be established during a correspondence update. Then based on the correspondences,

the ML-parameters are improved during a transformation update. However, the objective function in (5.15) extracted from (5.13) is fairly non-convex, due to the amount of parameters to update, the non-rigid nature of the transformation model and the presence of erroneous correspondences. The latter occurs mainly during the first iterations of the surface registration algorithm because of the severe difference in shape between the floating and the target surfaces after initialization. To cope with the non-convexity, we apply a deterministic annealing optimization, where the idea is to choose a control parameter changing the shape of the equivalent M-estimator (5.11) to construct a convex approximation of the objective function. The control parameter is slowly adjusted such that the objective function increasingly approximates the original non-convex estimation. Therefore, instead of treating the noise as an extra parameter to estimate, within the equivalent M-estimator, we replace it by a newly introduced deterministic parameter $\sigma^2 = T_\sigma$ similar to [71]. The annealing is accomplished by gradually reducing T_σ during the surface registration process. For a certain T_σ setting the correspondences and the transformation parameters are updated according to:

Correspondence-update: During this step a set of correspondences is established for the inlier points $\mathbf{r}_j^f \in \mathfrak{J}$ for $j = 1, \ldots, N^{\mathfrak{J}}$, with $N^{\mathfrak{J}}$ the number of inliers, according to:

$$\mathbf{r}_j^t = \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) - \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) \frac{\nabla \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))}{\|\nabla \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))\|} \tag{5.16}$$

Transformation-update: Based on the correspondences from the previous step the ML-parameters are to be updated according to (5.15) with $\sigma^2 = T_\sigma$ or equivalently:

$$\boldsymbol{\theta}^{(m+1)} = argmin_{\boldsymbol{\theta}} \left[\sum_{j \in \mathfrak{J}} \|\mathbf{r}_j^t - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\|^2 + 2T_\sigma \nu \ \mathrm{trace}(\Lambda^{\mathrm{T}} \Phi \Lambda) \right] \tag{5.17}$$

Note that the term $2T_\sigma \log \sqrt{\pi T_\sigma}$ is eliminated, due to its independency on $\boldsymbol{\theta} = \{C, \Lambda\}$. Solving (5.17) based on the corresponding points $S^f = \{\mathbf{r}_j^f \in \mathfrak{J}\}$ and $S^t = \{\mathbf{r}_j^t \in \mathfrak{J}\}$ is accomplished using the following closed form linear system:

$$\begin{pmatrix} \Phi + 2T_\sigma \nu I & S^f \\ {S^f}^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} \Lambda \\ C \end{pmatrix} = \begin{pmatrix} S^t \\ 0 \end{pmatrix} \tag{5.18}$$

where S^f and S^t are notated as $N^{\mathfrak{J}} \times 4$ matrices consisting of concatenated homogeneous points $\mathbf{r}_j^f$ and $\mathbf{r}_j^t$ for $j = 1 \ldots, N^{\mathfrak{J}}$ in every row. The size of the spline kernel Φ is now $N^{\mathfrak{J}} \times N^{\mathfrak{J}}$ and is only dependent on the amount of inlier points. Note that the transformation parameters are determined based on the original positions of the floating surface points, resulting in a single final

transformation after optimization instead of a set of consecutive transformations. After every transformation parameter the process (LV- and M-step) is repeated for a different setting of T_σ. Outliers are transformed as well, because an outlier can become an inlier in between different registration iterations and vice versa. Furthermore, after convergence missing data is recompleted by transforming the outlier points using the transformation based on the inliers.

Annealing: The deterministic control parameter T_σ is updated after every iteration or LV- and M-step execution according to a linear annealing scheme $T_\sigma^{(m+1)} = rT_\sigma^{(m)}$ with $0 < r < 1$ the annealing rate. Starting at $T_\sigma^{(0)} = T_\sigma^{init}$ the number of iterations is determined until $T_\sigma^{(m+1)}$ drops below a predefined value $T_\sigma^{final} < T_\sigma^{init}$. The higher T_σ the lower the confidence in the correspondences and based on (5.18) we see that the deformation of the floating surface is restricted. The lower T_σ the higher the confidence in the correspondences and the deformation of the floating surface is flexible. This is the key point behind the deterministic annealing optimization: the registration is performed using a coarse-to-fine matching strategy. First the global floating and target surface differences are eliminated and slowly, but deterministically, the local surface differences are eliminated. We believe that, if a person was asked to fit an elastic rubber mask of a face onto a solid sculpture of a different face, he or she would also start with a global adaption of the mask, while slowly but surely compensating for smaller and smaller differences, similar to the annealing used during optimization. Using the annealing scheme, the complete facial surface registration is described in pseudo-code as:

Initialize
$T_\sigma = T_\sigma^{init}$, $C = I$ and $\Lambda = 0$
procedure
while $T_\sigma \geq T_\sigma^{final}$ **do**
  LV-step
    Update inliers and outliers using (4.10)
  M-step
    Update correspondences using (5.16)
    Update transformation using (5.18)
  Annealing
    $T_\sigma = rT_\sigma$
end

To conclude this section we must note that the simplification of (5.15) extracted from (5.13), when having a combined (regular and random) latent variable is quite similar. Outlier points for which $z'_j = f(\mathbf{r}_j(\boldsymbol{\theta})) = 0$, with $f(\mathbf{r}_j(\boldsymbol{\theta}))$ according to (B.19), and for which $z''_j = b_j = 0$, with b_j according to (B.16), are neglected to determine $\mathfrak{J}$ and the matrix part $\Phi + 2T_\sigma \nu I$ in (5.18)

becomes $\Phi + 2T_\sigma \nu B$ with:

$$B = \begin{bmatrix} 1/b_1 & & \\ & \ddots & \\ & & 1/b_{N\mathfrak{Z}} \end{bmatrix} \tag{5.19}$$

Doing so, the spline smoothing is made individual point inlier-belief b_j dependent during transformation parameter updating. Points with a higher inlier-belief should be obeyed or interpolated more than points with lower inlier-belief allowing to be more approximated instead of interpolated in favor of smoothness.

5.3 Reference face

The choice of reference face, being the floating surface, is of importance. Ideally the reference face should not differ too much from the target faces in order to guarantee a proper non-rigid surface registration outcome. Starting to build a database, a specific face from the database can be selected as reference face, but it is not advisable to use for example the face of a young man as floating surface to register with the face of an old woman as target surface. Furthermore, a statistical model extracted from registered faces is biased towards the reference face. Ideally, the average face of the database is selected, due to its central position within the database, not differing too much from every face in the database and eliminating the model bias towards a specific face. However, the average face can only be determined based on registered faces, leading to a chicken and egg paradigm. To cope with this, multiple non-rigid registration runs throughout the whole database are performed. The first run, uses a specific face from the database, which is complete and for which a consistent uniform tessellation is obtained. After the non-rigid registration of every face in the database with the specific face the average face is computed and used as the reference face during a second run. This is repeated until the change in average face drops below a certain accuracy threshold.

The resulting reference face is shown in figure 5.1. The reference tessellation in figure 5.1(a) consists of 9327 vertices and 18311 triangles encoding the connectivity. In order to speed up calculations, not all vertices are used as floating surface points during registration. Instead, a subset of 1516 points is selected, shown in figure 5.1(b), such that more points are retained in areas with higher curvature information. After registration, the complete set of vertices with known connectivity is transformed using the transformation obtained with the subset of floating surface points. Moreover, additional points indicated on the reference face are automatically indicated on every face of the database using the obtained transformations. In figure 5.1(c), the reference face is depicted with indicated 52 face landmarks plus the nose tip using a distinctive color per landmark for visual validation of the non-rigid registration in the next section.

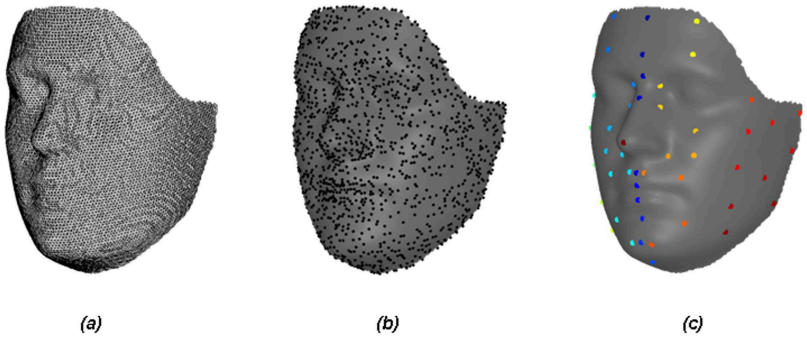


Figure 5.1: Illustration of the reference face, being the average face of the database. (a) The reference tessellation. (b) Shaded rendering with the points indicated used during non-rigid registration. (c) Same shaded rendering, with the 52 face landmarks indicated using a distinctive color for every landmark. This is shown to compare the landmark consistency registration results in figure 5.2(c) with the landmark positions of the reference face

5.4 Results

The registration strategy from the previous section while making use of the non-rigid registration method from section 5.2, is applied to our database consisting of 393 target faces. The choice of user-specific regularization ν in (5.15) is dependent on the amount of floating surface points and by experience and experiments $\nu = N^f = 1516$ is a good value. Doing so, an annealing scheme with $T_\sigma^{init} = 1$, $T_\sigma^{final} = T_\sigma^{init}/500$ and $r = 0.9$ is appropriate and is used.

Five non-rigid surface registration results after the last registration run with the reference face of figure 5.1 are depicted in figure 5.2. In figure 5.2(a) shaded renderings of the five original target surfaces, originating from different faces after facial acquisition (chapter 4), are shown. Note the incompleteness of the original faces due to missing data acquisition. The mesh tessellations and shaded renderings of the faces after non-rigid surface registration are illustrated in figure 5.2(b) and (c), respectively. Observe that the five results are represented by the same amount of points sharing the same connectivity, therefore being registered to each other besides the reference face as well. Furthermore, we see that only the facial part corresponding with the reference face is retained while missing parts are completed. Similar results are obtained for all the other faces in the database.

Based on the results we are particulary interested in the analysis of two different validation aspects: accuracy and consistency. The former aspect is discussed in section 5.4.1 and the latter is discussed in section 5.4.2

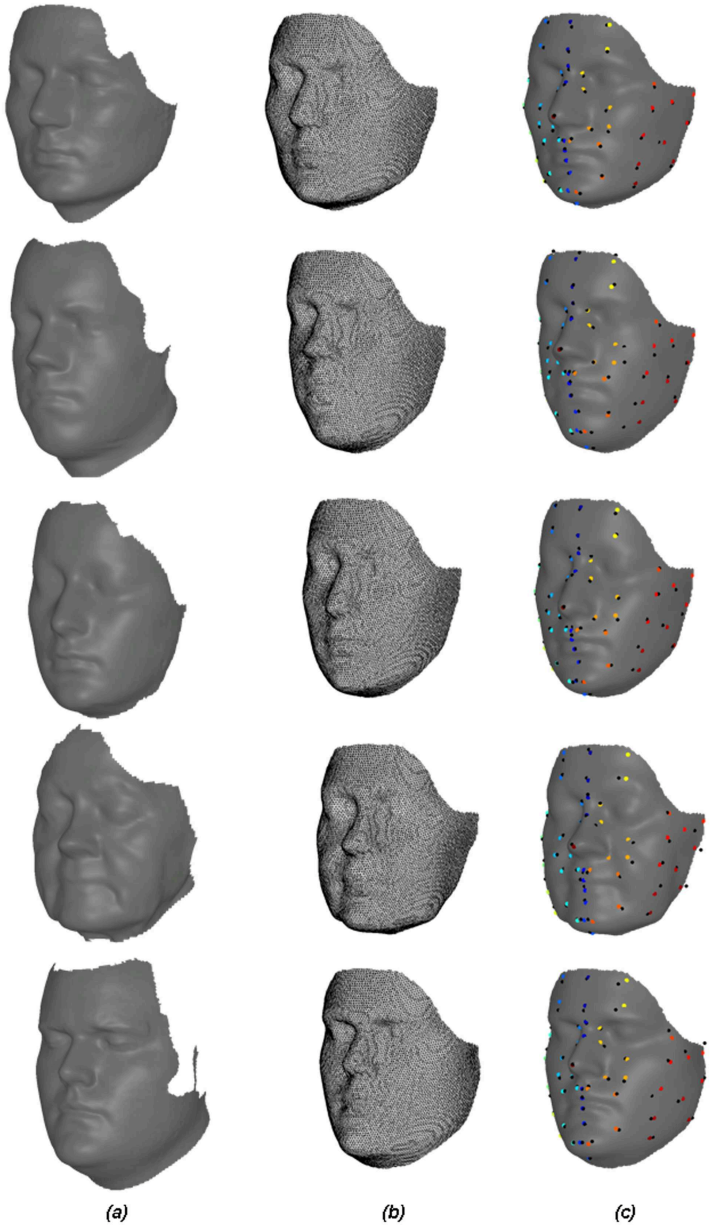


Figure 5.2: Five (Top to bottom) non-rigid surface registration results. (a) Original target surface . (b) The target surface tessellation after non-rigid registration. (c) A shaded rendering of the registration result with two sets of 53 indicated face landmarks. The colored face landmarks are the landmarks obtained by transforming the face landmarks of the reference face after registration. The black face landmarks are manually indicated face landmarks, using a blue eyeliner pencil during tissue-depth acquisition. Additionally the nose tip was indicated manually as well.

5.4.1 Accuracy

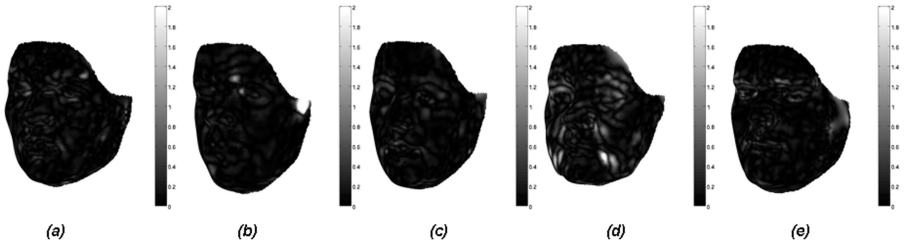


Figure 5.3: Accuracy analysis of the five results from figure 5.2. (a) - (e) correspond to top to bottom faces in 5.2. The error range is from 0 mm (black) to 2 mm (white)

The accuracy aspect analyzes how well the transformed floating surface describes the shape or geometry of the target surface and can for instance be obtained by computing the local surface errors $e_j = \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))$ using the VIS of the original target surface, which is shown in figure 5.3. It can be seen that the accuracy of the results is satisfying and usable in practice: the transformed floating surface is a accurate enough description of the target surface geometry. The only minor result is the cheek area of the fourth face (figure 5.3(d)), originating from an older woman. This is mainly because the reference face is quite young compared to the fourth face and because less floating surface points are sampled within the cheek area. The cheek area of the older woman contains more surface curvature than the reference face. This can be resolved by using more points in the cheek area, trading off computation speed for accuracy.

5.4.2 Consistency

The consistency aspect analyzes how well corresponding points on different faces are correct or consistent. The quality of the shape model is mainly dependent on this consistency aspect but is much harder to analyze or validate compared to the accuracy aspect. A first attempt is to visually inspect the consistency of certain key facial points like the 52 anatomical face landmarks at which tissue depths were measured (cf. appendix C) plus the nose tip depicted in figure 5.2(c). The colored face landmarks are obtained by transforming the landmarks of the reference face after registration and the color-code is according to the reference face in figure 5.1(c). It is visually seen that corresponding face landmarks are located at corresponding anatomical landmarks over the five faces and the reference face. However, a proper quantification of the consistency as such cannot be obtained. A second attempt is to compare the same face landmarks (colored) with the manually indicated face landmarks (black) during tissue-depth acquisition , using a blue eyeliner pencil. Additionally the nose tip is indicated manually as well. Visually it is seen that both black and

colored landmarks are very alike. However, we know that the consistency of the black landmarks themselves is not optimal due to the manual indication, being error-prone, and therefore cannot be used as ground truth for consistency quantification. A proper consistency quantification can be obtained by analyzing whether the knowledge of correspondences within a set of training faces is capable of explaining, or is consistent with the correspondences of a test face not part of the training set. This can be accomplished by using the training faces to build up a shape model and to register this model, being the floating surface now, with the surface of the test face. The model registration result can then be compared with the correspondences of the test face. Due to the fact that a shape model is to be constructed we postpone the consistency validation of the non-rigid registration to chapter 6 after we elaborated the statistical model construction and registration.

5.5 Conclusion

The last prerequisite for our craniofacial database was to obtain registered inter-subject facial surfaces to construct a statistical craniofacial model. Therefore, the 3D points with known connectivity of a carefully constructed generic reference face were mapped or registered non-rigidly onto the faces in the database. The facial surface registration methodology had to be robust against noise and had to be be able to cope with outliers due to partial overlap or missing data. Therefore we reformulated the registration problem within the robust surface registration framework of chapter 3. The reference face was considered to be the floating surface represented as a set of floating points, while a face from the database was considered to be the target surface represented with a VIS such that no sampling errors occurred. A proper non-rigid TPS based transformation model and similarity measure, based on the VIS target surface representation, were chosen and both were combined into a single objective function. To handle a vast amount of floating surface points, we used the VIF theory and algorithmic improvements to update and evaluate the transformation. Outliers were dealt with using an outlier detection function equal to the one used in the previous chapter. A deterministic annealing optimization strategy altering between inlier, point correspondence and transformation updates was applied to cope with the non-convex nature of the objective function. Some results were shown to perform a validation concerning accuracy and consistency. The former was easy to analyze and was acceptable. The latter was more difficult to analyze, but already some visual consistency results were shown and were promising.

Using the theoretical aspects given in this second part of the thesis, a facial database for craniofacial purposes can be built while minimizing the required amount of manual labor. In the next and final part of the thesis we acquire a database according to appendix C and chapters 4 and 5. Further details about the database and the construction of a statistical craniofacial model based on the database are given in chapter 6. Furthermore, additional

correspondence consistency validation is given as well based on the model. The proposed statistical craniofacial model together with alternative models are used in chapter 7 to perform craniofacial reconstructions based on a dry skull substrate, which is validated in chapter 7 as well.

Part III

Craniofacial modeling and reconstruction

Chapter 6

Statistical craniofacial modeling

6.1 Introduction

Within this third and final part of the thesis, the focus is upon the main application: statistical craniofacial modeling and reconstruction. The term statistical concerns both the modeling and the reconstruction methodology making the overall approach completely probabilistic. Given a skull specimen, the reconstruction step, elaborated in the next chapter, tries to find the facial solution that is most probable, according to a statistical craniofacial model. This chapter concentrates on the construction of the statistical craniofacial model capturing the relationship between skull and skin surfaces.

In order to eliminate template-related model bias and to minimize the unrealistic character of craniofacial reconstructions, we propose a flexible statistical craniofacial model constructed from a database of facial entries acquired following the procedures presented in chapter 4 and [1, 2] (appendix C), and co-registered following the non-rigid registration procedures of chapter 5. Every facial entry within the database consists of a facial shape (with inter-subject point correspondences), gray-value texture, tissue-depth at the 52 anatomical skin landmarks and property information (gender, age, BMI). A multi-variate Gaussian model is fit to he samples in the database by Principal Component Analysis (PCA) generating an average, being the *model template*, and a correlation-ranked set of principal components, being the *transformation model*, modeling the combined population-dependent variance and covariance of complete skin surface shape, gray-value texture, 52 anatomical face landmarks with tissue depths and property (age, BMI and gender) values, being the *model template information*. The craniofacial model template can be classified as a complete facial surface template from which a single most plausible reconstruction is obtained based on the analyzed knowledge of multiple faces. The incorporation of gray-value texture variation generates a more lifelike re-

construction based on a texture map which co-varies with facial geometry, property value and tissue depth information, leading to a more, but not too, specific texture map, compared to an average texture map. The creation of more specific reconstructions matching the subject properties estimated during an anthropological examination is obtained by modeling and removing facial variations originating from property differences in the facial database. The result is a property normalized craniofacial model.

The craniofacial model can be considered as an elastic mask of which the elasticity is defined as the statistically allowed correlated variation. By changing the statistical model parameters between the statistically determined boundaries, the mask is deformed in a face-specific way only. This mask flexibility constitutes a non-rigid statistical transformation model and can be used to solve different registration problems like the complete facial surface registration and the skull registration or reconstruction. The latter is elaborated in the next chapter. The former is similar to chapter 5 and is accomplished by using the average face of the database, being the model template, again as the floating surface, but the generic non-rigid TPS transformation used in chapter 5 is replaced by face-specific deformations extracted from facial covariance knowledge within the model. Basically, the incorporation of covariance knowledge between different facial information parts is the major difference of the statistical model proposed here, with existing models for craniofacial reconstruction. Eliminating subsequently covariance between different information parts (shape, thickness, ...) from the model leads to more traditional craniofacial models working with a single facial template in combination with generic transformation models (e.g TPS based).

The remainder of this chapter is organized as follows. We start with a short description and some remarks concerning the database in section 6.2. Based on the acquired and registered database the construction and completeness of the statistical model is given in section 6.3. Aspects of using a statistical model to solve the facial surface registration problem are given in 6.4, such that a further correspondence consistency validation is obtained. Model manipulation according to facial properties is elaborated in section 6.5.

6.2 Database

Using the methodologies described in chapters 4 and 5 of the second part concerning shape and texture information and described in [1, 2] (appendix C) concerning tissue depth information, a facial database is acquired and registered. A single facial entry has: (1) a facial surface in upright position described with dense corresponding points, (2) a gray-value converted texture for every point on the facial surface, (3) 52 specific anatomical landmarks or points on the facial surface, (4) 52 ultrasound soft-tissue depths measured within the facial landmarks and (5) a BMI, age and gender value. The reason for using gray-valued instead of RGB-valued texture information is twofold. Firstly, we believe that using gray-valued texture information, containing fewer

(unknown) details, is less dangerous to trigger the recognition process into the wrong direction. The texture information is always difficult to guess correctly based on the skull, such that the task of texture must be limited to: "giving an idea of lips, eyebrows, etc. without specific clues if they are not known". Secondly, the colored texture information is of low quality for the majority of the facial entries in the database due to a malfunctioning of the 3D acquisition camera.

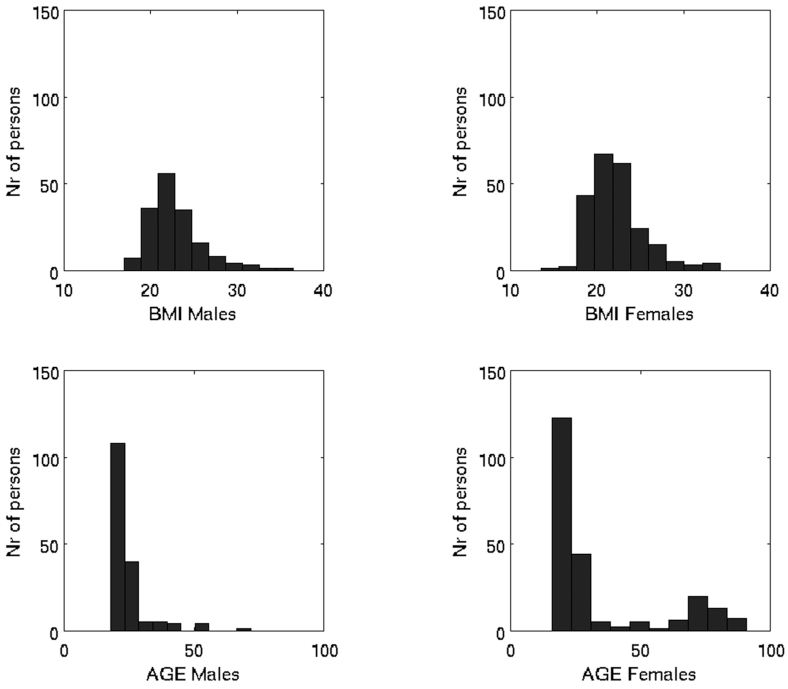


Figure 6.1: Database property distributions. BMI (top row) and AGE (bottom row) histograms for both males (left column) and females (right column).

Within the database we have a total of 393 persons of which 168 are male and 225 are female. BMI and age distributions of the database are depicted in figure 6.1. The minimum, maximum and average age are 18/72/24 and 16/91/34 for males and females respectively. The majority of the entries are young individuals, due to the fact that Leuven has a major student population. Furthermore, the majority of the older people are female. Note that this is not the most ideal age variation and distribution for statistical modeling and manipulation and will have to be kept in mind during section 6.5 concerning property manipulation. The minimum, maximum and average BMI are 16.9/36.5/22.8 and 16/34.2/22.2 for males and females respectively. The majority of the population is situated around 22-25, being considered normal: not to skinny or obese.

6.3 Statistical model

Based on the database described in the previous section a statistical craniofacial model, modeling the combined population-dependent variance and covariance of complete skin surface shape and texture, 52 anatomical face landmarks with tissue depths and property (age, BMI and gender) value information, is constructed and explored within this section. The model construction is accomplished using principal component analysis (PCA) after conversion of the facial entries to facial vector descriptions in section 6.3.1. Analysis of the model validity (accuracy and generalization) is given in section 6.3.2.

6.3.1 Construction

Based on the dense inter-subject correspondences obtained using the non-rigid surface registration procedure from chapter 5, the database can be represented using vector descriptions of the faces $\{\mathbf{F}_k | k = 1, \ldots, K\}$ with $K = 393$ the number of faces and with $\mathbf{F} = [\mathbf{F}^s, \mathbf{F}^g, \mathbf{F}^{lm}, \mathbf{F}^{td}, \mathbf{F}^p]^{\mathrm{T}}$ a column vectorial representation of a single face, with $^{\mathrm{T}}$ meaning the transpose of a vector, where $\mathbf{F}^s$ is the facial surface, $\mathbf{F}^g$ the gray value texture, $\mathbf{F}^{lm}$ the face landmark position, $\mathbf{F}^{td}$ the tissue depth and $\mathbf{F}^p$ the property information using the following vector representations:

$$\begin{aligned}
\mathbf{F}^s &= [r^s_{1x}, r^s_{1y}, r^s_{1z}, \ldots, r^s_{Nx}, r^s_{Ny}, r^s_{Nz}] \\
\mathbf{F}^g &= [g_1, \ldots, g_N] \\
\mathbf{F}^{lm} &= [r^{lm}_{1x}, r^{lm}_{1y}, r^{lm}_{1z}, \ldots, r^{lm}_{Lx}, r^{lm}_{Ly}, r^{lm}_{Lz}] \\
\mathbf{F}^{td} &= [d_1, \ldots, d_L] \\
\mathbf{F}^p &= [b, a, s]
\end{aligned} \tag{6.1}$$

N is the number of 3D points on a facial surface (9327 in our case) and L the number of face landmarks (52 in our case). g_i is the gray value of the point $\mathbf{r}^s_i$ on the facial surface, while d_i is the tissue depth value of the $\mathbf{r}^{lm}_i$ face landmark point. Both $b(BMI)$ and $a(age)$ have continuous values, while the $s(sex/gender)$ values are initially discrete (-1 for females and +1 for males) but are converted to continuous values according to appendix D in order to differentiate between the male and female alikeness of faces. The more positive/negative a gender value the more male/female a face appears to be. A value of zero corresponds to a genderless looking face. As mentioned in chapter 5, some facial entries are incomplete due to the presence of hair during 3D acquisition, resulting into missing shape and texture information values within the corresponding facial vector descriptions which are to be completed before a proper PCA can be conducted. Missing shape information refilling is obtained trough the non-rigid facial surface registration method explained in chapter 5. Corresponding missing texture information refilling is accomplished using a probabilistic missing data PCA procedure described in [100].

Ensuring corresponding point coordinates, gray-values, tissue depths and property values on the same row of column vector descriptions from different

faces, a face can be represented as a single point in a $(3\times N+N+3\times L+L+3)$ dimensional *face space*. Within this face space new acceptable faces can be expressed as linear combinations of the faces in the database:

$$M(\mathbf{w}) = \sum_{k=1}^{K} w_k \mathbf{F}_k \tag{6.2}$$

with reasonable limits on the range of the parameter values w_k. A craniofacial model can then be defined as the set of faces $M(\mathbf{w}) = [M^s(\mathbf{w}), M^g(\mathbf{w}), M^{lm}(\mathbf{w}), M^{td}(\mathbf{w}), M^p(\mathbf{w})]^{\mathrm{T}}$ parameterized by the coefficients $\mathbf{w} = (w_1, \ldots, w_K)$. However, such a high dimensional face space which is discretely encoded using facial samples is rather impractical to work with. Therefore a probability distribution is associated to this linear face space by fitting a multivariate normal distribution to the data set of K faces, resulting into a continuous parametrization and a dimension reduction of the face space. The multivariate normal distribution fitting is accomplished using a Principal Component Analysis (PCA) of the covariance matrix X constructed from the mean normalized faces $\breve{\mathbf{F}}_k = \mathbf{F}_k - \overline{\mathbf{F}}$ with $\overline{\mathbf{F}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{F}_k$ the average face (shown in figure 6.2 (a)):

$$X = \tfrac{1}{K} F F^T \quad \text{with} \quad F = [\breve{\mathbf{F}}_1 | \breve{\mathbf{F}}_2 | \ldots | \breve{\mathbf{F}}_K] \tag{6.3}$$

PCA computes the eigenvectors $U = \{\mathbf{U}_k | k = 1, \ldots, K-1\}$ and eigenvalues $\Sigma = \{\sigma_k^2 | k = 1, \ldots, K-1\}$ of the covariance matrix X, which can be achieved using a Singular Value Decomposition of the covariance matrix [101]. PCA performs essentially a basis transformation to an orthogonal coordinate system (eigenspace) spanned by the eigenvectors or principal components U of the covariance matrix in descending order according to their associated eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_{K-1}^2$, representing the variances of the data along each eigenvector. In theory a total amount of $K-1$ eigenvectors and eigenvalues can be extracted based on K faces in the database. However the lower components contain very small and useless variance information (see figure 6.2(b)) and can be discarded, leading to a dimension reduction. In practice a number $D = 41$ smaller than $K - 1 = 392$ components are retained modeling 98% of the total variance in the database. The remaining 2% variance is regarded as noise and is discarded. The statistical craniofacial model is then defined as the set of faces $M(\mathbf{c}) = [M(\mathbf{c})^s, M(\mathbf{c})^g, M(\mathbf{c})^{lm}, M(\mathbf{c})^{td}, M(\mathbf{c})^p]^T$ parameterized by the coefficients $\mathbf{c} = (c_1, \ldots, c_D)$ according to:

$$M(\mathbf{c}) = \overline{\mathbf{F}} + \sum_{k=1}^{D} \mathbf{U}_k c_k = \overline{\mathbf{F}} + U\mathbf{c} \tag{6.4}$$

where $\mathbf{U}_k = [\mathbf{U}_k^s, \mathbf{U}_k^g, \mathbf{U}_k^{lm}, \mathbf{U}_k^{td}, \mathbf{U}_k^p]^{\mathrm{T}}$, because different rows of an eigenvector influence different aspects of the facial information. $\mathbf{U}^s$ affects the facial surface information, $\mathbf{U}^g$ influences the gray value texture information, $\mathbf{U}^{lm}$ changes the face landmark position information, $\mathbf{U}^{td}$ affects the tissue depth information and $\mathbf{U}^p$ is related with the property information. In addition PCA provides an estimate of the probability density within the face space in terms of the Mahalanobis-distance MD to the origin or average face of the

eigenspace:

$$p(\mathbf{c}) \sim MD(\mathbf{c}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\sum_{k=1}^{D}\frac{c_k^2}{\sigma_k^2}\right) \tag{6.5}$$

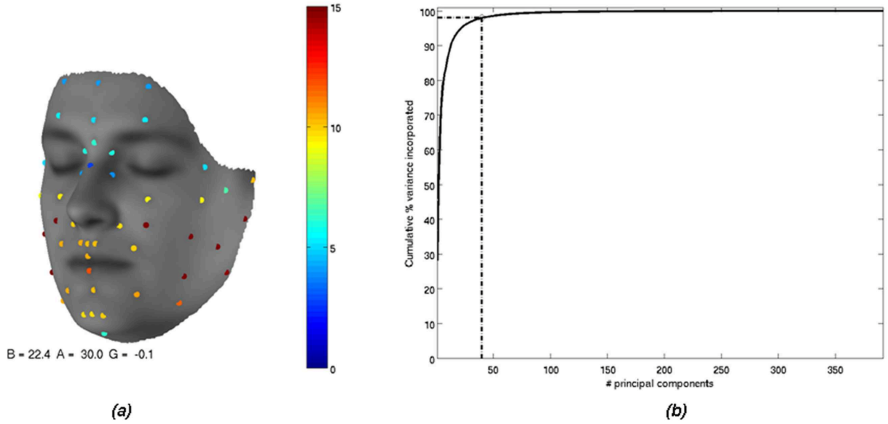


Figure 6.2: (a) the average face of the database. Tissue-depths are visualized within the appropriate facial landmarks using a color-code ranging from 0 (blue) to 15 (red) mm. The average BMI = 22.4, age = 30 and gender = -0.1 (slightly female but mainly genderless). (b) cumulative percentage variance curve (solid line). Horizontal axis: the number of principal components incorporated. Vertical axis: percentage of the total variance captured by the number of components. 98% variance is captured using only 41 of the 392 principal components (dotted lines)

Using (6.4), every facial entry in the database can be represented using the parametric representation of the statistical model. Indeed, instead of using a vector description as in (6.1), the facial entry can now be modeled as the sum of the averaged face and a weighted linear combination of the principal modes of variation. The parameters describing the facial entry are the weights $\mathbf{c}$ in this linear combination. A probability value can be associated to each set of parameters or facial entry using (6.5), with maximal probability for the average face. By altering the parameters, in between statistically determined boundaries ($-3\sigma_k \leq c_k \leq 3\sigma_k$) learned from the database, faces spanning a face space similar to the database can be generated. The average face can be altered in a face-specific way by changing the coefficients or parameters $\mathbf{c}$. In figure 6.3 the effects on the average face by changing the model parameters associated to the first three modes of variation are shown. The first mode of variation seems to characterize the difference between young males and old females. The second mode characterizes the facial differences between compressed and elongated faces along an anterior-posterior axis. The third mode of variation incorporates besides a type of facial structure difference also

a difference in BMI. These principal modes of variation can be considered being face-specific deformations, because they were learned based on faces, while staying within statistical boundaries a facial plausible outcome is guaranteed.

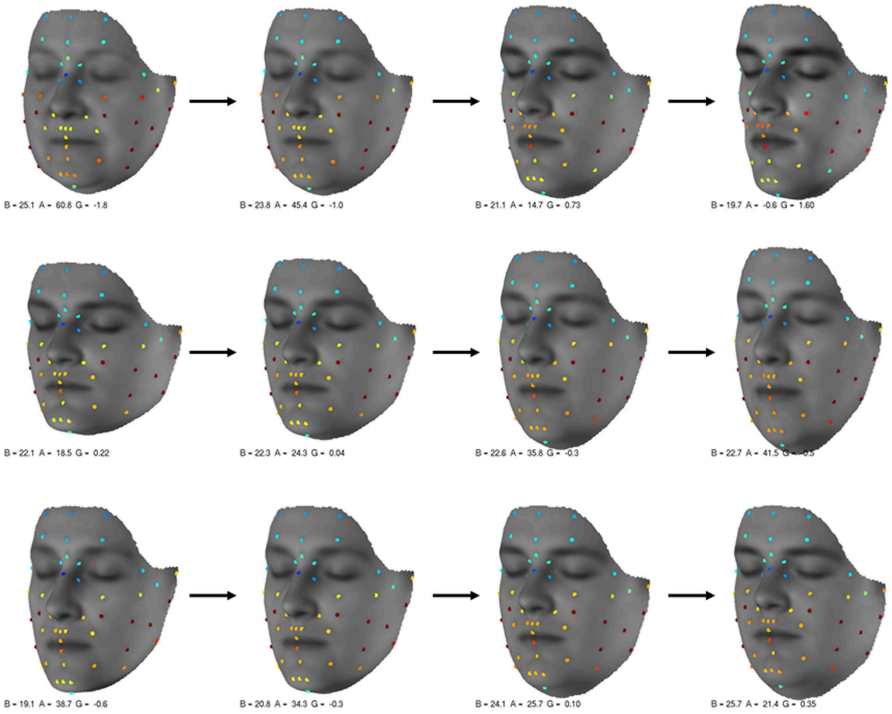


Figure 6.3: Visualization of the first (top row), second (middle row) and third (bottom row) principal modes of variation or face-specific deformations. From left to right appropriate principal component weights c_k go from $-3\sigma_k$ over $-1.5\sigma_k$ and $1.5\sigma_k$ to $3\sigma_k$. Tissue depths are visualized similar

Alternating or updating the average face using restricted weighted combinations of principal components as in figure 6.3 is the key point of using a statistical model for skull reconstruction purposes. Within the craniofacial model context and nomenclature defined in chapter 2, we see, based on (6.1)and (6.4), that the average face is the craniofacial model template incorporating shape, texture, tissue depth and property values, being the model template information and that the set of principal components is the craniofacial transformation model consisting of face-specific deformations. The craniofacial model template can be classified as a complete facial surface template from which a single most plausible reconstruction is obtained based on the analyzed knowledge of multiple faces.

6.3.2 Model validity

By altering the parameters, in between statistically determined boundaries in figure 6.3 it is already preliminary shown that the model is capable of generating new and plausible faces. In this section we try to analyze the model sufficiency or whether the model is capable of creating valid example faces within a reasonable accuracy. For this, samples of the database can be used to examine how well a wider population can be modeled. However, one has to keep in mind that the test is only valid for the database on which it is run. Two sufficiency tests adapted from [102] are performed .

Model accuracy

The first test analyzes how well facial entries, within the set used to create the model or training set, are described in terms of accuracy using models capturing different amounts of total database variance expressed in percentage. The question to be answered is: How many percentage variance is sufficient? In a first instance the number of principal components needed to describe a certain amount of variance is determined to construct model descriptions of the facial entries within the training set. Subsequently, the accuracy of the shape is computed by calculating the Root-Mean-Squared-Error (rmse) of the distances between corresponding points of the model description and an original facial entry of the training set. The accuracy of the thickness and texture information is computed as the average absolute difference between the model description and the facial entry thickness and texture information. The property value accuracy is the absolute difference between both. This is done for every facial entry in the database and the accuracy results for the different facial information parts are averaged and given in table 6.1. As expected the accuracy error decreases as the percentage of modeled variation increases, becoming zero using 100% database variance. However, the number of principal components increases. Already mentioned in previous section, typically 98% is used in literature, at which level we see that the number of required principal components is limited to 41 instead of 392 and at which the accuracy errors are acceptable.

Model generalization

The second test analyzes how well facial entries, not part of the training set or test faces, are described in terms of accuracy using models based on different amounts of training faces. Having enough training data we expect the model to be able of describing unseen or test faces quite accurately or to generalize well. The question then becomes: how many training faces are sufficient? Table 6.1, gives the target accuracy for a specific percentage of variance (98%) and having enough training faces the same accuracy results are expected for the test faces. For different amounts of training data we select test faces not used to build up the model. The model parameters of a test face **F** are obtained

Perc. Var.	90	95	98	99	99.9	100
Nr PC	13	22	41	62	176	392
Shape	1.08	0.76	0.47	0.33	0.09	0.00
Thickness	1.79	1.72	1.58	1.32	0.56	0.00
Texture	0.062	0.061	0.059	0.057	0.039	0.00
BMI	1.93	1.82	1.59	1.43	0.45	0.00
Age	7.5	5.7	3.5	2.2	0.2	0.00
Gender	0.38	0.26	0.09	0.02	0.01	0.00

Table 6.1: Accuracy evolution by increasing the total database variance captured by the model. Nr PC are the amount of principal components required to model the desired variance in percentage. Shape and thickness are expressed in mm, texture ranges from 0 to 1 and has no specific unit. BMI, age (in years) and gender are the errors concerning the properties.

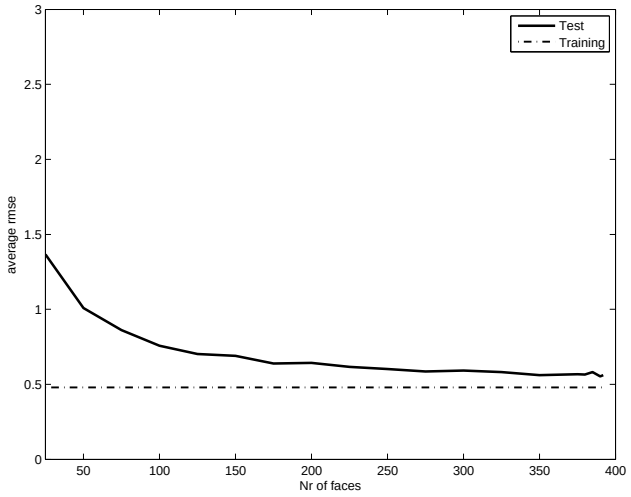


Figure 6.4: Shape accuracy evolution for different amounts of training data (solid curve). Target shape accuracy (dotted line) Horizontal axis: nr of training faces, Vertical axis: averaged rmse.

as:

$$\mathbf{c} = U^{\mathrm{T}}(\mathbf{F} - \overline{\mathbf{F}}) \tag{6.6}$$

These model parameters are then used in (6.4) to create the model description of the test face and the same accuracy calculations from the previous test are performed. This is done for 200 test faces per amount of training faces were both test faces and training faces are randomly selected. The accuracy results of all the 200 test faces for a certain amount of training data are averaged. Figure 6.4 depicts the shape accuracy evolution with increasing amount of training faces. Similar results were obtained for the other facial information parts. It can be seen that the accuracy errors approach their in-training-set values as more training faces are added. Despite the small difference at the end of the curve (using 392 training faces) we can conclude that the amount of training faces in our database seems sufficient when modeling 98% of the training data variance. The only beneficial effect of adding more examples is a better represented remaining 2% of the population.

6.4 Model based facial surface registration

As mentioned earlier, the key point of using the statistical model , being a craniofacial model for skull reconstruction purposes, is the ability to deform or transform the average face non-rigidly using restricted weighted combinations of principal components. More generally, the statistical model can be placed within the registration framework of chapter 3 to tackle other non-rigid registration problems, like the facial surface registration, as well. The average face, being the model template, is the floating surface. The target surface depends on the registration problem and can either be another facial surface, during facial surface registration, or a skull surface, during skull registration. A proper similarity criterium depending on the registration problem is determined based on the model template information. The transformation model, being statistical, is the key point of using a statistical model for registration purposes.

Applying the statistical model for skull registration is elaborated in the next chapter. In this section we use the statistical model to tackle the facial surface registration problem similar to chapter 5. Doing so, correspondence consistency validation can be obtained. First we elaborate the required registration components in section 6.4.1 after which the results are given in section 6.4.2.

6.4.1 Registration framework

Quite a few registration components are similar to the ones defined in section 5.2. The same similarity measure and outlier process are applied. The floating surface, being the average face, is represented by a set of points $S^f = \{\mathbf{r}_j^f | j = 1 \ldots, N^f\}$ which are either the complete set or a subset of the average facial surface information $S^f \subseteq \overline{\mathbf{F}}^s$ or $\{\mathbf{r}_j^f = \overline{\mathbf{r}}_j^s | j = 1 \ldots, N^f\}$ with $N^f \leq N$.

The target surface, being a facial surface, is represented using a VIS $S^t = \mathfrak{d}^t(\mathbf{r})$ augmented with a VBF $\mathfrak{b}^t(\mathbf{r})$ in case target facial shape information is missing. The main two differences with section 5.2 are the transformation model, being statistical based instead of TPS based, and the optimization, using an analytical gradient descent approach instead of an ICP approach with deterministic annealing.

Combining (6.4) and (6.5), we have a statistical non-rigid transformation model parameterized as a weighted sum of face specific deformations or principal components (6.4) incorporating a statistical regularization (6.5). The regularization includes a statistical facial plausibility according to a face space distribution. Therefore, the prior transformation model distribution will favor transformation parameters generating more plausible facial instances. Furthermore, the statistical transformation model ensures that the floating surface is deformed in a face-specific way only. However, the current transformation model is not capable of rotating and translating the floating surface, due to the fact that faces of the database are rigidly aligned with the average face prior to PCA computation and as such no rigid transformation knowledge is incorporated. The PCA based transformation of a floating surface point $\mathbf{r}_j^f$ including a rigid body can be defined as:

$$\mathbf{r}_j^{f*} = \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) = R\mathbf{r}_j^f + \mathbf{t} + U_j^s\mathbf{c} \tag{6.7}$$

with R a rotation matrix (4.6) and $\mathbf{t}$ a translation vector (4.7) both defined according to three basis axes with their origin placed in the geometric mean of the average face ($\mathbf{o} = \sum_j^N \overline{\mathbf{r}}_j^s/N$). $\mathbf{c} = \{c_k|k = 1, \ldots, D\}$ are the weights or statistical model parameters for the principal components in the eigenvector matrix U and U_j^s is the row of the eigenvector matrix influencing $\mathbf{r}_j^f$. The complete set of transformation parameters becomes $\boldsymbol{\theta} = [\omega_x, \omega_y, \omega_z, t_x, t_y, t_z, \mathbf{c}]$. Because no regularization is needed for the rigid transformation body the prior parameter distribution $p(\boldsymbol{\theta})$ or transformation model likelihood reduces to $p(\mathbf{c})$ equal to (6.5), resulting in

$$\|\mathfrak{L}(\boldsymbol{\theta})\|^2 = \frac{1}{2}\sum_{k=1}^{D}\frac{c_k^2}{\sigma_k^2} \tag{6.8}$$

,regularizing the deformation towards plausible facial solutions. The prior knowledge about the registration problem expressed in the deformation through this negative log model likelihood is based on the variances σ_k^2 incorporated in a facial database. In other words, the registration is performed keeping in mind (6.8) and based on (6.7) the knowledge of multiple faces from a database.

Excluding the outlier process for simplification purposes, without loss of generality, the objective function defined in (4.17) using (6.7) and (6.8) becomes:

$$\mathfrak{E}(\Theta) = \sum_{j=1}^{N^f}\left[\frac{1}{2\sigma^2}\mathfrak{d}^t(R\mathbf{r}_j^f + \mathbf{t} + U_j^s\mathbf{c})^2 + \log\sqrt{2\pi}\sigma\right] + \nu\frac{1}{2}\sum_{k=1}^{D}\frac{c_k^2}{\sigma_k^2} \tag{6.9}$$

with $\Theta = \{\boldsymbol{\theta}, \sigma\}$ incorporating the noise standard deviation besides the transformation parameters. Note that the number of parameters to estimate is $D+7$ with $D = 41$ for a statistical model capturing 98% variance. Furthermore, the number of parameters is independent on the amount of floating surface points in contrast to the TPS based transformation parameters. Therefore we opted for a gradient descent optimization approach to minimize (6.9). Doing so, no explicit correspondences are required and the non-convexity of the objective function is well behaved such that no deterministic annealing on the noise parameter is needed. Instead the noise parameter is treated as an extra parameter to update or estimate.

The transformation parameters are updated by executing one step of Newton's method for object function optimization, which is a gradient descent method based on the gradient $\nabla\mathfrak{E}(\boldsymbol{\theta})$ and the Hessian $\mathfrak{H}(\boldsymbol{\theta})$ with ($\mathfrak{H}_{ij} = \partial^2\mathfrak{E}(\boldsymbol{\theta})/\partial\theta_i\partial\theta_j$) resulting into:

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - s\mathfrak{H}(\boldsymbol{\theta})^{-1}\nabla\mathfrak{E}(\boldsymbol{\theta}) \tag{6.10}$$

The gradient is calculated analytically $\nabla\mathfrak{E}(\boldsymbol{\theta}) = [\partial\mathfrak{E}/\partial\theta_1, \ldots, \partial\mathfrak{E}/\partial\theta_{D+6}]^{\mathrm{T}}$ based on the first derivatives of (6.9) to every parameter in $\boldsymbol{\theta}$. $\mathfrak{H}(\boldsymbol{\theta})$ is based on the second derivatives and is computed by numerical differentiation from the analytically calculated first derivatives. The gradient $\nabla\mathfrak{E}(\boldsymbol{\theta})$ defines the search direction for the parameter update. The Hessian captures information about an appropriate order of magnitude of updates in each parameter and is updated every 50 e.g. iterations. The optimal step-size s is found by line-minimization. The update of σ in case of gaussian noise and no outlier process is:

$$\sigma^{(m+1)} = \sqrt{\frac{1}{N^f}\sum_{j=1}^{N^f}\mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}^{(m)}))^2} \tag{6.11}$$

This updating process is iteratively repeated until convergence occurs after which a final set of transformation parameters is obtained. Corresponding points for the floating surface points are then obtained using (6.7). Because the eigenvectors are extracted from the covariance matrix X, they describe covarying changes between the different facial surface parts or points. Changing a subset of facial surface points, being the floating points, results in a covarying change of other facial surface points $\{\mathbf{r}_i^f = \overline{\mathbf{r}}_i^s | i = 1\ldots, N - N^f\} \neq \{\mathbf{r}_j^f | j = 1\ldots, N^f\}$ not used during registration. Therefore, corresponding points for the latter after registration can be obtained using (6.7) as well with the same transformation parameters applied on the relevant rows U_i^s of the eigenvector matrix. In fact, correspondence knowledge in the floating surface points after registration is statistically interpolated based on the covariance knowledge incorporated in the eigenvectors.

Having elaborated the framework for facial surface registration based on the statistical model, it is interesting to compare the statistical transformation defined in (6.7) with the TPS based transformation. To pronounce similarities

the TPS transformation defined in (5.2) is rewritten into:

$$\mathbf{r}_j^{f*} = \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) = R\mathbf{r}_j^f + \mathbf{t} + \boldsymbol{\phi}(\mathbf{r}_j^f)\Lambda \tag{6.12}$$

by using a rigid instead of an affine subspace through the incorporation of scale and skew into the non-rigid subspace. It is clearly observed that both (6.7) and (6.12) consist of a rigid and a non-rigid subspace. The eigenvectors and the spline kernel are responsible for the non-rigid deformation contribution, but the difference is that the eigenvector matrix is built from the covariance matrix expressing the internal relationship of inter-subject floating surface points from multiple faces while the spline kernel expresses the internal relationship of floating surface points originating from a single reference face. Using a TPS based generic transformation the prior knowledge or model-likelihood used during registration consist of a single reference face or template in combination with a certain generic smoothness constraint, regularizing the generic deformation of the template. Correspondence information in the floating surface points after registration is mathematically smooth but generically interpolated to establish additional correspondences. Using a PCA based or face-specific transformation the prior knowledge or model-likelihood used during registration consists of multiple reference faces expressed by the eigenvectors and is regularized according to the eigenvalues, generating facial plausible solutions. Correspondence information in the floating surface points after registration is mathematically statistical but face-specifically interpolated to establish additional correspondences. The difference between the PCA and TPS based transformations and interpolations is depicted in figure 2.9 for a facial nose tip manipulation.

Alternatively, we can state that the statistical transformation model uses dense correspondence knowledge from a facial training set to establish dense correspondences with a new unseen face. This can be used to analyze the correspondence consistency within the database, which is done in the following section.

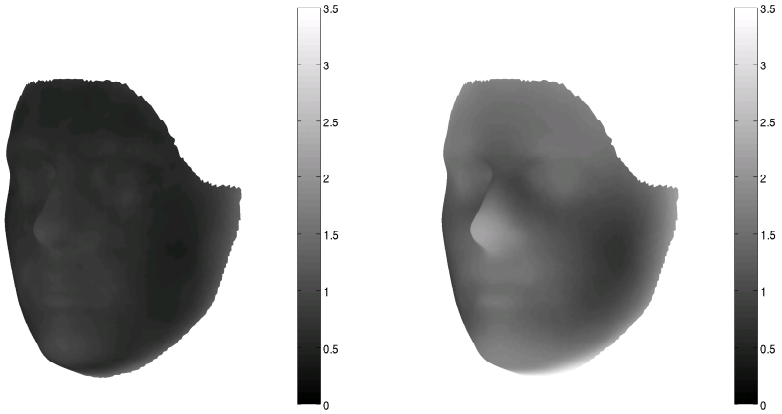
6.4.2 Results

Consistency validation

The quality of the statistical model concerning the facial shape information is related to the established dense inter-subject correspondences. The most important prerequisite for model building is the consistent determination of the correspondences in between inter-subject surfaces of the database, which is not straightforward to validate due the lack of proper ground truth information. As mentioned in chapter 5, a proper consistency quantification can be obtained by analyzing whether the knowledge of correspondences within a set of training faces is capable of explaining or is consistent with the correspondences of a test face not part of the training set. Using the statistical model machinery from this chapter we are able to conduct such a test. Recall from

the previous section that the eigenvector matrix contains inter-subject correspondence structural relationship knowledge being used during model based facial surface registration. In other words, correspondence knowledge from a training set is contained within the statistical model and can be used to validate whether the correspondences established on a test face , not part of the training set, using the TPS based registration framework (cf. section 5.2) are consistent with the correspondences within the training set using the statistical model based registration framework (cf section 6.4.1) applied on the test face.

To perform the consistency test a leave-one-out strategy is conducted. Every facial entry from the database is lifted from the database in turn, being the test face with known TPS based correspondences, while every other face is used to build up the model according to section 6.3. Sequently the model is registered with the test face according to section 6.4.2 to establish statistical model based correspondences. The same subset of 1516 floating surface points used for the TPS based facial surface registration of chapter 5 and visualized in figure 5.1(b) are used to perform the model based registration. Similarly, other corresponding points are obtained after the registration by evaluating the final transformation in the facial surface points of the average face not used during registration. The user-specific amount of regularization during model based registration is set equal to one $\nu = 1$, making the amount of regularization equal to the true statistical boundaries of the facial population. Increasing the value of ν would favor more averaged facial registration solutions restricting flexibility while decreasing ν would allow to obtain facial implausible solutions according to our database. After model based facial surface registration dense TPS based $\{\mathbf{r}_{kj}^{TPS} | j = 1, \ldots, N\}$ and model based $\{\mathbf{r}_{kj}^{PCA} | j = 1, \ldots, N\}$ point correspondences on the k'th test face are inherently known and are compared by calculating the euclidian distances $d_{kj} = \|\mathbf{r}_{kj}^{TPS} - \mathbf{r}_{kj}^{PCA}\|$, being local correspondence differences or errors, in between both sets of points. Based on these local errors a single $rmse_k = \sqrt{\sum_{j=1}^{N} d_{kj}/N}$ for the k'th face is computed. This is done for every face of the database $k = 1, \ldots, K$, such that an averaged rmse is obtained $rmse = \sum_{k=1}^{K} rmse_k/K$. Furthermore, a single $rmse_j = \sqrt{\sum_{k=1}^{K} d_{kj}/K}$ of local correspondence differences is obtained as well and is depicted in figure 6.5(a) using the facial geometry of the average face. Doing so, a spatial distribution of the point correspondence differences is created and visualized. Having a sub-millimeter global $rmse = 0.72mm$ and observing the local rmse in figure 6.5(a), we can conclude that the TPS based correspondences are accurately described using the statistical model based correspondences. Equivalently, we can conclude that the analyzed or statistical knowledge of correspondences within a training set is capable of finding the correspondences of an unseen test face such that the correspondences obtained in chapter 5 are considered to be consistent throughout the database.



(a) (b)

Figure 6.5: Model based registration results. (a) rmse of local correspondence differences using the same subset of floating surface points defined in section 5.3 for consistency validation. (b) rmse of local correspondence differences using the 52 face landmarks to analyze the co-varying facial surface results. The gray-color code ranges from 0 (black) to 3.5 mm (white)

Reconstruction preview

The model based registration framework in section 6.4.1 is tailored for the registration of facial surfaces originating from different individuals. The main application of the statistical model is craniofacial reconstruction elaborated in chapter 7. However, the methodology for craniofacial reconstruction is quite similar, except for a few differences. One of the main differences is the set of floating surface points used during registration. Due to the fact that skull relating information, being tissue depths, is limited too the 52 face landmarks, the floating surface point set consist of these sparse set of landmarks instead of dense facial surface points. Similar too section 6.4.1, because the eigenvectors are extracted from the covariance matrix X incorporating different facial information parts, they describe co-varying changes between the different information parts. Changing the facial landmarks and tissue depths e.g. results in a co-varying change of facial surface. For the craniofacial recosntruction problem it is interesting to examine the statistical influence of the 52 landmarks on the remaining facial surface. Therefore we conducted a second but very similar test to analyze the strength of the co-varying knowledge between the sparse set of landmarks and the remaining facial surface points. The same test setup and analysis is used as before except that the floating surface points from the previous test are replaced by the sparse set of 52 face landmarks. The global $rmse = 1.38mm$ and the local rmse results are depicted in figure 6.5(b). The performance is lower than the first test using 1516 floating surface

points, which is quite logical. However, keeping in mind the limited amount of evidence (52 floating surface points) used during registration, we can conclude that the co-varying power between the facial surface points encoded in the eigenvectors of the statistical model is quite strong. Equivalently, the statistical interpolation of the facial surface information based on the face landmark information after registration is within a desirable accuracy.

Besides the amount of floating surface points used during registration, Another difference between statistical model based facial surface and skull registration is the requirement for model property manipulation prior to registration for the latter, which is explained in the following section.

6.5 Facial property normalization

Within the craniofacial work-flow defined in chapter 2, the anthropological examination generates a set of skull properties, which are considered to be hard constraints to guarantee a proper recognition result. Therefore the craniofacial model is to be adapted accordingly. Traditionally, a subpopulation, having similar properties, of the database is selected and is used to build a craniofacial model. However, besides related issues given in section **??**, we see in section 6.2 that our database is not particulary suited to do this. Quite a few subpopulations have limited facial samples, resulting in inferior and non-flexible statistical models extracted from those subpopulations. Furthermore, facial variations not originating from differences in properties are thrown away. Instead we opt to learn and to remove or eliminate facial variations originating from property differences within the database while retaining all the other facial variations. The advantage is that under-sampled subpopulations can be represented by the interpolating nature of the learning process. After the elimination process, a property normalized statistical craniofacial model is obtained having enough flexibility, while the skull properties are obeyed as a hard constraint prior to skull registration. Section 6.5.1 explains the learning process, while section 6.5.2 concentrates on the removing process.

6.5.1 Learning

Having a database of different sub-populations, it is interesting to learn deformations originating from differences in property values between different sub-populations to simulate changes in age, BMI and gender on faces. In figure 6.3 we already saw that the first principal component incorporated a change in age and gender, while the third component incorporated a change in BMI. However using the first component in order to manipulate a face in terms of age/gender would also cause a simultaneous change in gender/age which is not practical for ageing or changing gender only simulations respectively. Using the third component to change the BMI property of a face would also cause a simultaneous change in face geometry not originating from a change in BMI, which is not useful for a change in BMI only simulation. Furthermore,

other components besides the first and the third may incorporate changes in properties. In this section we explain the extraction of independent property related variations and deformations originating from differences in property values while the fundamental geometry of faces is retained.

Expressing the original data in terms of principal components (6.4) is in fact a description of the data in terms of variances and covariances (6.3). The goal is to find proper directions in the principal component space related to property variations $\mathbf{F}^p = [b, a, s]$. The extraction of property dependent variations from the statistical model is obtained by finding three orthogonal basis vectors in the coordinate system spanned by the eigenvectors, describing a separate independent change of b, a and s. Since this can be achieved by an infinite number of linear combinations of the eigenvectors, the linear combination that has the smallest Mahalanobis-distance MD (most probable) to the origin of the eigenspace is searched for ensuring minimal shape variation and therefore retaining the fundamental geometry. According to [103] this can be solved by a constrained least-square optimization. Let $\mathbf{v}_b$, $\mathbf{v}_a$ and $\mathbf{v}_s$ be the three unknown basis vectors, defined in a D-dimensional eigenspace, describing a unit change of BMI, age and gender respectively. The Mahalanobis-distance MD of these three vectors is given by:

$$MD(\mathbf{v}_x) = \sum_{k=1}^{D} \frac{v_{xk}^2}{\sigma_k^2} \quad \text{for} \quad x \in \{b, a, s\} \tag{6.13}$$

The weights v_{xk} of the linear combination discribing the unit vectors are similar to the parameters c_k describing faces in 6.4. Taking into account that b, a and s values depend only on the last three rows of the eigenvector matrix U, the submatrix U^p consisting of the last three rows of U can be used. Minimization of the Mahalanobis-distance MD subject to the constraint of independent BMI, age and gender change by one unit is done solving the following Lagrange functional L:

$$\begin{gathered} L(\mathbf{v}_x, \mathbf{l}_x) = \sum_{k=1}^{D} \frac{v_{xk}^2}{\sigma_k^2} - \mathbf{l}_x^{\mathrm{T}}[U^p \mathbf{v}_x - \mathbf{e}_x] \\ x \in \{b, a, s\} \\ \mathbf{e}_b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_a = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_s = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{gathered} \tag{6.14}$$

The vectors $\mathbf{l}_x$ for $x \in \{b, a, s\}$ are the Lagrange multipliers. The vectors $\mathbf{e}_x$ for $x \in \{b, a, s\}$ are three unit vectors perpendicular onto each other, therefore being independent. The Lagrange functional $L(\mathbf{v}_x, \mathbf{l}_x)$ is solved by calculating the derivatives with respect to all the variables $\mathbf{l}_x$ and $\mathbf{v}_x$, resulting into the following two linear equations:

$$\begin{aligned} 2A^{-1}V &= {U^p}^{\mathrm{T}} Q \\ U^p V &= I \end{aligned} \tag{6.15}$$

where,

$$A = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_D^2 \end{bmatrix} \tag{6.16}$$

$$V = \begin{bmatrix} \mathbf{v}_b & \mathbf{v}_a & \mathbf{v}_s \end{bmatrix} \qquad Q = \begin{bmatrix} \mathbf{l}_b & \mathbf{l}_a & \mathbf{l}_s \end{bmatrix}$$

The three basis vectors in V are then given as:

$$V = AU^{p^{\mathrm{T}}}[U^p A U^{p^{\mathrm{T}}}]^{-1} \tag{6.17}$$

While $\mathbf{v}_b$ describes a change of BMI by one unit with constant age and gender having minimal MD and therefore generating the most plausible minimal facial shape variation, $\mathbf{v}_a$ and $\mathbf{v}_s$ alter *age* and *gender* correspondingly.

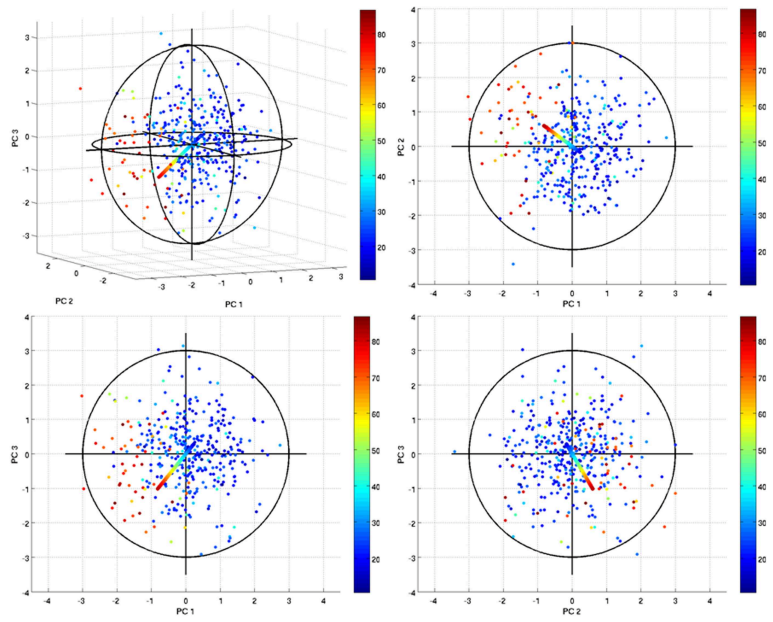
A visualization of the age basis vector $\mathbf{v}_a$ using a normalized scattered data plot of the facial samples in the database according to the first three principal components is given in figure 6.6(a). Using the color-codes, it is observed that the vector $\mathbf{v}_a$ indeed describes a direction in which the age property value of the faces changes accordingly. Additionally, we determined another age basis vector $\mathbf{v'}_a$, without the BMI and gender independency constraints encoded in $\mathbf{e}_x$ for $x \in \{b, a, s\}$ and is visualized in 6.6(b). Although, the alternative vector also visually describes a change in age, it is observed that the contribution of the first principal component to $\mathbf{v'}_a$ is bigger than $\mathbf{v}_a$. Because the first principal component incorporates a gender change besides an age change as well (due to our database content cf. section 6.2), using $\mathbf{v'}_a$ instead of $\mathbf{v}_a$ would result in a gender change during ageing simulations, which is not wanted. Similar results are obtained for the BMI vector $\mathbf{v}_b$ and the gender vector $\mathbf{v}_s$.

6.5.2 Removing

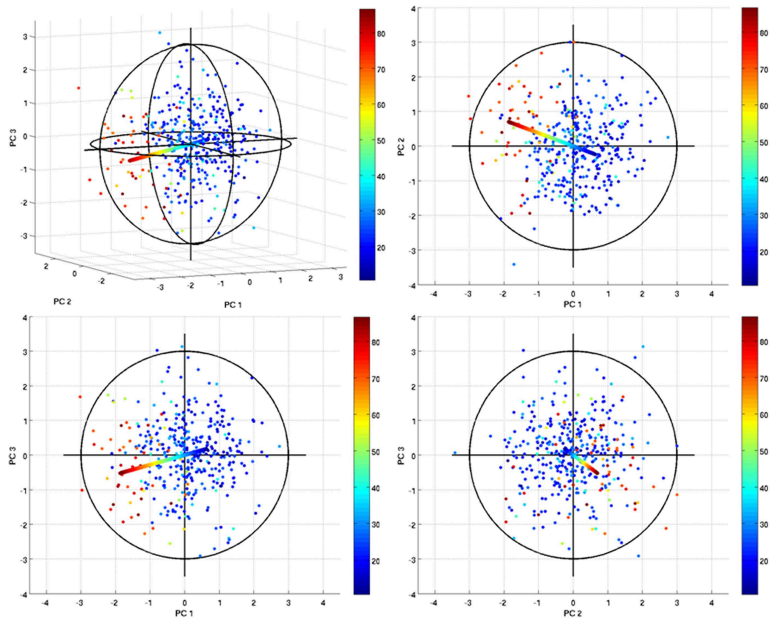
Having learned the property basis vectors or directions, properties of a face can be altered by adding a certain amount of these basis vectors to their model parameter descriptions, generating appropriate property manipulations. The effects caused by adding an amount of these basis vectors to the average face in figure 6.2(a) are shown in figure 6.7.

Manipulating or normalizing faces within the database, according to a given set of skull property values $[b_s, a_s, s_s]$, is the key point to removing facial variation due to property differences from the statistical craniofacial model. Given a facial sample of the database with property values $[b, a, s]$, the basis vectors $\mathbf{v}_x$, weighted by appropriate differences $\triangle x = x_s - x$ for $x \in \{b, a, s\}$, are subtracted from the facial model parameters $\mathbf{c}$ of the facial entry:

$$\tilde{\mathbf{c}} = \mathbf{c} - V \begin{bmatrix} \triangle b \\ \triangle a \\ \triangle s \end{bmatrix} \tag{6.18}$$



(a)



(b)

Figure 6.6: Visualization of the two age basis vectors (colored lines) $\mathbf{v}_a$ (a) and $\mathbf{v'}_a$ (b) .Top left, scattered data plot of faces from the database plotted as 3D points using normalized facial model parameters of the first three principal components $\{c_k/\sigma_k | k = 1, \ldots, 3\}$. The color values correspond to the age value of the faces. Top right, bottom left and right, similar but a 2D plot instead of a 3D plot against the first and second, first and third, second and third principal components respectively.

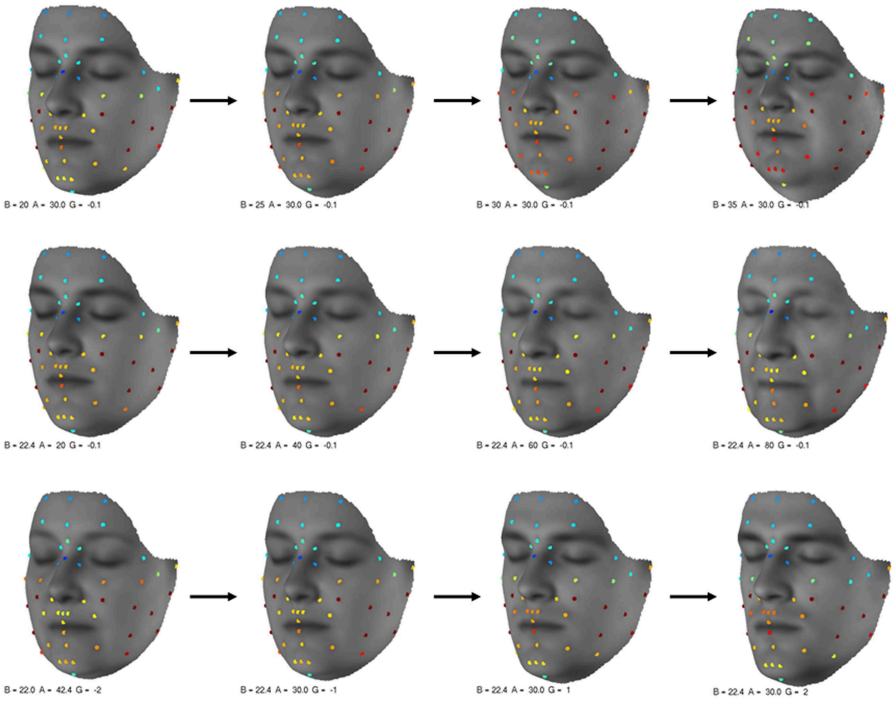


Figure 6.7: Illustration of property value changes applied onto the average face. Top row, BMI simulation going from 20 over 25 and 30 to 35. Middle row, Age manipulation going from 20 over 40 and 60 to 80 years old. Bottom row, Gender value change going from -2 over -1 and 1 to 2.

such that a property normalized facial entry is obtained according to:

$$\tilde{\mathbf{F}} = \overline{\mathbf{F}} + U\tilde{\mathbf{c}} \tag{6.19}$$

This is done for every facial sample in the database $\{\mathbf{F}_k | k = 1, \ldots, K\}$. As such, a new database is obtained which is centered around and only contains the given property values of the skull specimen. Every face in the database has now the same property values equal to the given values. The remaining variability in this *property value normalized* population is smaller than the original population and can be calculated by applying a new PCA to the normalized set of instances $\{\tilde{\mathbf{F}}_k | k = 1, \ldots, K\}$. The result is a new property value normalized statistical model:

$$\tilde{M}(\mathbf{c}) = \overline{\tilde{\mathbf{F}}} + \sum_{k=1}^{D-3} \tilde{\mathbf{U}}_k c_k = \overline{\tilde{\mathbf{F}}} + \tilde{U}\mathbf{c} \tag{6.20}$$

with

$$\overline{\tilde{\mathbf{F}}} = \frac{1}{K} \sum_{k=1}^{K} \tilde{\mathbf{F}}_k \tag{6.21}$$

and

$$\tilde{\mathbf{U}}_k = [\tilde{\mathbf{U}}_k^s, \tilde{\mathbf{U}}_k^g, \tilde{\mathbf{U}}_k^{lm}, \tilde{\mathbf{U}}_k^{td}]^T \tag{6.22}$$

There is no more eigenvector part $\tilde{\mathbf{U}}_k^p$ influencing the property information, because the variance and covariance of property information within the normalized database is zero. The new eigenvector or principal component basis $\tilde{U}$ models the face-specific differences within a certain sub-population according to the property values of the given skull specimen. Doing so, a set of given skull property values for a particular target skull can be imposed as a hard constraint prior to skull registration.

6.6 Conclusion

In this chapter we presented a statistical craniofacial model for skull reconstruction purposes, modeling the combined population dependent variance and covariance of skin surface shape, texture, tissue depths and properties. The model was constructed from a database of 393 individuals, using principal component analysis, generating an average face, being the model template, combined with a set of face-specific deformations or principal components, being the transformation model. The model capacity was examined for in-training and unseen faces and appeared to be sufficient in terms of incorporated percentage variance and number of training samples.

The statistical model, being an elastic mask, was embedded within the registration framework of chapter 3 to tackle a model-based facial surface registration problem and was compared to the TPS based approach in chapter 5. Doing so, a further correspondence consistency validation was performed with satisfying results. Furthermore a slight preview concerning skull instead of facial surface registration was given by using the sparse face landmarks instead of dense surface points as floating surface points. This showed an appropriate strength of the co-varying facial information knowledge incorporated in the eigenvector matrix to solve skull registrations based on a limited amount of registration evidence.

An additional prerequisite for a model being a valid craniofacial model was the ability to normalize or alter the model based on a given skull property set obtained from an anthropological examination. Instead of selecting a subpopulation from the database we opted to learn and to remove facial variations from the model originating from differences in property values. The result was a flexible (containing enough variation) but property normalized statistical craniofacial model such that skull properties are obeyed as a hard constraint prior to skull registration.

The reason for our model being statistical is because it is constructed based on the covariance matrix of the facial training data. Therefore analyzed knowledge of multiple faces is incorporated within the model and is used during skull reconstruction which is explained in the following chapter. Basically, the incorporation of covariance knowledge between different facial information parts is the major difference of the proposed statistical model with existing models

for craniofacial reconstruction. Eliminating subsequently covariance between different information parts from the model leads to more traditional craniofacial models working with a single facial template in combination with generic transformation models (e.g TPS based). Therefore, in the next chapter, we present three alternative craniofacial models by subsequently eliminating covariance information in order to compare our model with more traditional models. Independent of the craniofacial model choice a proper skull registration framework similar to section 6.4.1 using appropriate skull representations is elaborated in the next chapter to tackle the skull registration.

Chapter 7

Statistical craniofacial reconstruction

7.1 Introduction

Using the robust statistical surface registration framework, presented in chapter 3, this chapter discusses the final step in the CFR work-flow (section 2.2): reconstruction of the most plausible face from a given skull specimen. The choice of craniofacial model reflects our prior knowledge on the reconstruction problem, for which several choices exist. If the statistical facial model of the previous chapter is used, then the most plausible face is defined according to a facial distribution incorporating the knowledge of multiple faces. Furthermore the complete approach, both model and registration, is probabilistic. However, other (more traditional and non-statistical) models can be used as well by changing the most plausible face definition accordingly, e.g. based on the knowledge of a single face. Similarly, several choices of skull representations to represent the skull data exist as well. The only prerequisite is being compatible with the craniofacial model template. However, one has to keep in mind that a virtual copy of the skull is never anexact copy and that additional noise and outliers are incorporated during skull representation build up.

Within the registration framework context and nomenclature, CFR consists of a floating and target surface, transformation model, similarity measure and a robust objective function to optimize. Here, the target surface is the skull surface. The craniofacial model (model template and transformation model) determines the floating surface. The similarity measure is defined based on the choice of skull representation, which is almost inevitably corrupted by noise and outliers. Skull registration is then obtained by constructing and optimizing a robust objective function combining a transformation model and similarity measure, while being robust against noise and outliers.

This chapter elaborates and combines the different components needed for craniofacial reconstruction. We start with a general statistical based registra-

tion framework for craniofacial reconstruction in section 7.2, which is applicable to a wide range of craniofacial models and skull representations. Next, we give details of the full statistical craniofacial model from the previous chapter in section 7.3. Additionally, three alternative models are derived from the proposed model by progressively eliminating incorporated covariance knowledge. Doing so, we mimic several traditional craniofacial models while casting them in the registration framework as well. Then, we propose an implicit skull representation making use of variational implicit surfaces in section 7.4 in order to obtain a completely automatic reconstruction procedure. Additionally, we construct a more traditional craniometric skull representation as well, leading to a semi-automatic reconstruction procedure. Both skull representations are straightforwardly inserted into the registration framework. Results for the different craniofacial models and skull representations embedded within the statistical registration framework are given in section 7.5, while trying to perform a proper craniofacial reconstruction validation.

7.2 Craniofacial reconstruction framework

Basically, craniofacial reconstruction (CFR) can be formulated as presented in chapter 3 by selecting the appropriate components and outlier handling. In this section we start from the craniofacial reconstruction formulation and convert it into the surface registration formulation of section 3.3 in chapter 3. The abstract statistical craniofacial reconstruction formulation is: *What is the most plausible or probable face (Recon) , according to a craniofacial model (Model), given the (erroneous) skull data (Skull) ?* Mathematically, this is similar to maximizing the following conditional probability:

$$Recon = \text{argmax}_{\ Model} p(Model|Skull) \tag{7.1}$$

In order to solve this question a proper craniofacial model and skull representation is required. Suppose we have a complete skull surface, which is the *target surface* S^t in the registration context, represented such that the distance $\mathfrak{D}(S^t, \mathbf{r})$ for points $\mathbf{r}$ to the skull surface can be obtained. Furthermore, suppose we have a craniofacial model, consisting of a model template, being the *floating surface* S^f, and a transformation model $\mathfrak{T}(\mathbf{r}, \boldsymbol{\theta})$ with regularization functional $\|\mathfrak{L}(\boldsymbol{\theta})\|^2$. Finally, suppose that the complete set or a subset of model template points, being the *floating surface points* $\{\mathbf{r}_j^f | j = 1, \ldots, N^f\}$, contains distance to skull model template information, being the *predefined distances* $d_j(\boldsymbol{\theta})$, such that a link, being the *individual point similarities* $\mathfrak{s}_j(\boldsymbol{\theta}) = \mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) - d_j(\boldsymbol{\theta})$, between both the model and the skull can be established at these points. Then the statistical craniofacial reconstruction formulation can be converted into a probabilistic surface registration formulation: *What are the most probable parameters, according to a prior transformation model, transforming the floating surface towards noisy target surface data while increasing matching quality.* Using (3.19) and (3.20) the mathematical

formulation in (7.1) is converted into the following MAP formulation:

$$Recon = \text{argmin}_{\boldsymbol{\theta}} \left(-\log p(S^t|\boldsymbol{\theta}) - \log p(\boldsymbol{\theta}) \right) \tag{7.2}$$

The second term is the model-likelihood as defined in (3.15), which is completely dependent on the choice of craniofacial model (cf section 7.3). The first term is the complete data-likelihood as is dependent on the individual point similarities extracted from the model template information combined with the skull representation (cf section 7.4). Dealing with noise or small errors within the individual point similarities the data-likelihood is modeled using a gaussian probability density function (3.22), such that the complete set of parameters to estimate includes a noise standard deviation $\Theta = \{\boldsymbol{\theta}, \sigma\}$. Additionally, outlier robustness is accomplished using a complete-process modeling for the data-likelihood combining an inlier- and outlier-process, through the introduction of an outlier latent variable map $\mathbf{z}$. Having a complete skull surface, outlier presence is mainly due to the skull digitization process (CT scanning with amalgam teeth fillings for example) and the skull representation build up (incorrectly indicated craniometric landmarks, for example). Due to the limited knowledge about the outlier generation process and due to the non-binary nature of the outliers we propose to use a random latent variable map. Having an incomplete skull, additional outliers are present due to partial overlap and the random latent variable can then be combined with a regular latent variable with strong deterministic knowledge as explained in section 3.3.3. For simplicity reasons, without loss of generality, we assume a complete skull surface is at hand. Incorporating outlier handling, the first term in (7.2) is then:

$$p(S^t|\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(S^t, \mathbf{z}|\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(S^t|\mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z}) \tag{7.3}$$

with $p(\mathbf{z})$ defined in (B.2) using fractional a-priori outlier knowledge and $p(S^t|\mathbf{z}, \boldsymbol{\theta})$ being the complete-process defined in (B.1). Similar to section 3.3.3, using an uniform outlier process distribution $p^o(\mathfrak{s}_j) = \delta$ and taking the negative log likelihoods of (7.2), we obtain the following objective function:

$$\mathfrak{E}(\Theta, \mathbf{z}) = \mathfrak{S}(\Theta, \mathbf{z}) + \nu \|\mathfrak{L}(\theta)\|^2 \tag{7.4}$$

with

$$\mathfrak{S}(\Theta, \mathbf{z}) = \sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j)) \tag{7.5}$$

where

$$h(x, \alpha(z), \beta(z)) = \alpha(z)x^2 + \beta(z) \tag{7.6}$$

and

$$\alpha(z) = \tfrac{1}{2\sigma^2} z \quad \beta(z) = -(1-z)\log\delta + z \log\sqrt{2\pi}\sigma \tag{7.7}$$

The optimal solution is found through a joint parameter estimation, leading to a dual updating procedure. Due to the probabilistic nature of the outlier map

being a random latent variable map, the joint estimation can be obtained using an Expectation-Maximization (EM) optimization. A sequence of parameter estimates $\{\Theta^{(m)}|m = 0, 1, \ldots\}$ is obtained by alternating the following two steps:

E-step: Taking into account (7.3), on the $(m+1)^{th}$ iteration, the latent variables are eliminated from the complete negative log data-likelihood $\mathfrak{S}(\Theta, \mathbf{z})$ trough out-integration, by computing the conditional expectation $E[x|y]$ of the complete negative log data-likelihood:

$$\mathfrak{S}(\Theta) = E[\mathfrak{S}(\Theta, \mathbf{z})|\mathbf{z}] \tag{7.8}$$

w.r.t the posterior $p(\mathbf{z}|S^t, \Theta)$. In other words, during the E-step the values of the latent variables z_j in the negative complete log data-likelihood (B.5) are replaced by the posterior $p(z_j|\mathfrak{s}_j, \Theta)$ values. Following the Mean Field strategy and using the easily interpretable prior outlier parameter κ of section 3.3.3, concerning the random latent variable choice, we obtain the following equivalent M-estimator:

$$\mathfrak{S}(\Theta) = \sum_{j=1}^{N^f} \rho(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha, \beta) \tag{7.9}$$

with

$$\rho(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha, \beta) = h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j^{(m+1)}), \beta(z_j^{(m+1)})) \tag{7.10}$$

where

$$z_j^{(m+1)} = b_j(\widehat{\Theta}^{(m)}) = \frac{p(\mathfrak{s}_j|\widehat{\Theta}^{(m)})}{p(\mathfrak{s}_j|\widehat{\Theta}^{(m)}) + \lambda} \tag{7.11}$$

and where λ is related to κ according to (B.18). Note that z_j is also dependent on the noise standard deviation σ estimate besides the model parameters $\boldsymbol{\theta}$ of the previous iteration. The estimated values of the latent outlier variables at the $(m+1)^{th}$ iteration reflect the inlier-belief or confidence for a certain individual point similarity, which is taken into account during the M-step.

M-step The ML parameters $\widehat{\Theta}^{(m+1)}$ are then updated minimizing the energy in (7.4), where (7.5) is replaced by (7.9) with fixed latent variable values:

$$\mathfrak{E}(\Theta) = \mathfrak{S}(\Theta) + \nu\|\mathfrak{L}(\theta)\|^2 \tag{7.12}$$

The transformation parameters are updated by executing one step of Newton's method for object function optimization, which is a gradient descent method based on the gradient $\nabla\mathfrak{E}(\boldsymbol{\theta})$ and the Hessian $\mathfrak{H}(\boldsymbol{\theta})$ with $(\mathfrak{H}_{ij} = \partial^2\mathfrak{E}(\boldsymbol{\theta})/\partial\theta_i\partial\theta_j)$ resulting into:

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - s\mathfrak{H}(\boldsymbol{\theta})^{-1}\nabla\mathfrak{E}(\boldsymbol{\theta}) \tag{7.13}$$

The gradient is calculated analytically based on the first derivatives of (7.12) to every parameter in $\boldsymbol{\theta}$. $\mathfrak{H}(\boldsymbol{\theta})$ is based on the second derivatives and is computed by numerical differentiation from the analytically calculated first derivatives.

The gradient $\nabla\mathfrak{E}(\boldsymbol{\theta})$ defines the search direction for the parameter update. The Hessian captures information about an appropriate order of magnitude of updates in each parameter and is updated every 50 e.g. iterations. The optimal step-size s is found by line-minimization. The update of σ in case of gaussian noise and the outlier process is:

$$\sigma^{(m+1)} = \sqrt{\frac{\sum_{j=1}^{Nf} z_j^{(m+1)} \mathfrak{s}_j(\widehat{\boldsymbol{\theta}}^{(m)})^2}{\sum_{j=1}^{Nf} z_j^{(m+1)}}} \tag{7.14}$$

This EM updating process is iteratively repeated until convergence occurs after which a final set of transformation parameters is obtained. These can then be used to transform other model template surface points, besides the landmark points used during registration, and other information (e.g. texture) in order to acquire a complete craniofacial reconstruction. Having presented the general probabilistic based skull registration framework, specific craniofacial model and skull representation choices are given in the following two sections of this chapter, starting with the craniofacial models.

7.3 Craniofacial models

The choice of craniofacial model mainly influences the transformation model and therefore the model likelihood term in (7.2) and the transformation parameters $\boldsymbol{\theta}$ to estimate during the M-step. Furthermore, the model template influences the amount of floating surface points used during registration and the predefined distances, depending on the face-skull relating information contained within the database from which the model is deducted. Having a database containing both complete skull and facial surfaces, the floating surface points could be a set of skull surface points, while the predefined distances are equal to zero ensuring the model skull surface to be aligned with the target skull surface after registration. Having a database similar to ours, the relationship between the model template and the skull is defined in $L = 52$ face landmarks with ultrasound tissue depths. Therefore the floating surface points are considered to be the anatomical facial landmarks, while the predefined distances are equal to the tissue depths. The registration purpose is to reposition the facial landmarks at a distance to the skull equal to the according tissue depths.

In this section, we define four craniofacial model choices based on our database. The relationship between the models and the skull is always defined in L face landmarks, being the floating surface points , combined with tissue depths. Necessary details concerning the transformation model, predefined distances and registration result interpolation of these four models are given. Starting from the statistical model in chapter 6, being our primary choice, the three other models are derived by eliminating subsequently facial covariance knowledge, resulting in two statistical-based and two TPS-based craniofacial

models. The former two contain a statistical based transformation model, due to the presence of appropriate covariance knowledge. The latter two contain a TPS based transformation model, due to the lack of appropriate covariance knowledge. We start with the statistical based choices after which the TPS based choices are presented and discussed.

7.3.1 Statistical based

First model (M_0)

The first craniofacial model M_0 , being our primary choice, is the statistical model $M(\mathbf{c})$ defined in the previous chapter after a property normalization $\tilde{M}(\mathbf{c}) = M_0$. This property normalization is done to obey a given set of skull properties prior to CFR and is feasible thanks to the covariance information between property values and other facial information parts (cf. section 6.5). During registration a set of model parameters $\mathbf{c}$ is estimated alternating the normalized average face landmark locations, being the floating surface points $\{\mathbf{r}_j^f = \overline{\tilde{\mathbf{r}}}_j^{lm} | j = 1, \ldots, L\}$, accordingly. Additionally, similar to section 6.4, a rigid transformation is incorporated to align the model instance with the skull in the same coordinate system. Therefore, the floating surface point transformation for the first model (M_0) is defined as:

$$\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) = R\overline{\tilde{\mathbf{r}}}_j^{lm} + \mathbf{t} + \tilde{U}_j^{lm}\mathbf{c} \tag{7.15}$$

with $\tilde{U}_j^{lm}$ three rows of the eigenvector matrix influencing the coordinates of $\overline{\tilde{\mathbf{r}}}_j^{lm}$ and $\boldsymbol{\theta} = [\omega_x, \omega_y, \omega_z, t_x, t_y, t_z, \{c_k | k = 1, \ldots, D-3\}]$. D are the number of principal components retained, modeling 98% variance, while $D-3$ are the number of principal components retained after property normalization. Using (7.15) the face landmarks are transformed in a face-specific way only due to the inter-landmark covariance knowledge coded in $\tilde{U}_j^{lm}$. Furthermore, thanks to the incorporation of thickness covariance, the predefined distances, being normalized tissue depths, are dependent on the model parameters and are alternated as well:

$$d_j(\boldsymbol{\theta}) = \overline{\tilde{d}}_j + \tilde{U}_j^{td}\mathbf{c} \tag{7.16}$$

with $\tilde{U}_j^{td}$ a single row of the eigenvector matrix affecting $\overline{\tilde{d}}_j$. Because no regularization is needed for rigid transformations the prior parameter distribution $p(\boldsymbol{\theta})$ reduces to $p(\mathbf{c})$ resulting in the following model regularization:

$$\|\mathfrak{L}(\boldsymbol{\theta})\|^2 = \frac{1}{2}\sum_{k=1}^{D-3}\frac{c_k^2}{\tilde{\sigma}_k^2} \tag{7.17}$$

regularizing the deformation towards plausible facial solutions. The prior knowledge about the reconstruction problem expressed in the deformation through this log model-likelihood is based on the variances $\tilde{\sigma}_k^2$ incorporated in a facial database. In other words, the reconstruction is performed based

on ((7.15)and (7.16)) and keeping in mind (7.17) the covariance of multiple reference face landmarks combined with tissue depths from a database. Alternatively, facial plausibility is expressed in terms of facial surface combined with facial tissue depth probability. The craniofacial model compared to which the most plausible facial solution is searched for during registration is in fact a facial distribution containing combined facial surface and thickness information.

Finally, thanks to the covariance knowledge between landmarks with tissue depths and other facial surface points with gray texture values captured in $\tilde{U}^s$ and $\tilde{U}^g$, the co-varying textured facial surface estimate or reconstruction is obtained using:

$$\begin{aligned} F^s_{est} &= \{R\overline{\tilde{\mathbf{r}}}^s_i + \mathbf{t} + \tilde{U}^s_i\mathbf{c}|i = 1,\ldots,N\} \\ F^g_{est} &= \{\overline{\tilde{g}}_i + \tilde{U}^g_i\mathbf{c}|i = 1,\ldots,N\} \end{aligned} \tag{7.18}$$

with N the number of facial template surface points. As a result the registration defined in the facial landmarks only is statistically interpolated and extrapolated in order to obtain a final textured reconstruction result.

Second model (M_1)

The second craniofacial model M_1 is quite similar to the previous one, with the only exception that covariance knowledge between thicknesses and facial surface shape/geometry is eliminated. Suppose no database with combined facial surface and tissue-depth information is available, but instead two separate databases for facial surface, $\{\mathbf{F}^A_k|k = 1,\ldots,K_A\}$ with $\mathbf{F}^A = [\mathbf{F}^s, \mathbf{F}^g, \mathbf{F}^{lm}, \mathbf{F}^p]^T$ and tissue depths $\{\mathbf{F}^B_k|k = 1,\ldots,K_B\}$ with $\mathbf{F}^B = [\mathbf{F}^{td}, \mathbf{F}^p]^T$ are used, based on the vector definitions in (6.1). In this case no covariance between thicknesses and facial geometry can be established and used during reconstruction, instead two separate statistical models $(M^A_1(\mathbf{c}^A), M^B_1(\mathbf{c}^B))$ must be constructed and applied, constituting the second craniofacial model. Similar to model M_0, both databases can be normalized towards a given set of skull properties based on the property covariance resulting in two separate normalized models $(\tilde{M}^A_1(\mathbf{c}^A), \tilde{M}^B_1(\mathbf{c}^B)) = M_1$. However, there is no statistical connection anymore between $\tilde{U}^A$ and $\tilde{U}^B$, so that varying thicknesses/facial surfaces do not cause co-varying facial surface/thickness respectively. The floating surface point transformation is dependent on $\tilde{M}^A_1(\mathbf{c}^A)$ and is defined as:

$$\mathfrak{T}(\mathbf{r}^f_j, \boldsymbol{\theta}) = R\overline{\tilde{\mathbf{r}}}^{lm}_j + \mathbf{t} + \tilde{U}^{Alm}_j\mathbf{c}^A \tag{7.19}$$

and the predefined distances, being normalized tissue depths of the thickness database, are dependent on $\tilde{M}^B_1(\mathbf{c}^B)$:

$$d_j(\boldsymbol{\theta}) = \overline{\tilde{d}}_j + \tilde{U}^{Btd}_j\mathbf{c}^B \tag{7.20}$$

The complete set of parameters is now $\boldsymbol{\theta} = [\omega_x, \omega_y, \omega_z, t_x, t_y, t_z, \mathbf{c}^A, \mathbf{c}^B]$, while the model-likelihood and the regularization become:

$$\begin{aligned} p(\theta) &= p(\mathbf{c}^A)p(\mathbf{c}^B) \\ \|\mathfrak{L}(\theta)\|^2 &= \tfrac{1}{2}\sum_{k=1}^{D^A-3}\tfrac{(c_k^A)^2}{(\tilde{\sigma}_k^A)^2} + \tfrac{1}{2}\sum_{k=1}^{D^B-3}\tfrac{(c_k^B)^2}{(\tilde{\sigma}_k^B)^2} \end{aligned} \tag{7.21}$$

,regularizing the deformation towards facial plausible solutions. The difference with M_0 is that plausibility is not expressed in terms of thicknesses combined with facial surfaces, but both independently from each other. In other words it is possible to generate a plausible facial landmark and a plausible tissue-depth configuration, but when combined and evaluated within M_0 it is not guaranteed to be plausible. Alternatively, the craniofacial model compared to which the most plausible facial solution is searched for during registration is in fact two separate independent distributions, one containing facial surface information and one containing thickness information.

Finally, thanks to the covariance between landmarks (without tissue depths) and complete facial surfaces with gray texture values captured in $\tilde{U}^{As}$ and $\tilde{U}^{Ag}$, the co-varying textured facial surface estimate or reconstruction can be obtained. Just replace $\tilde{U}^s$,$\tilde{U}^g$ and $\mathbf{c}$ in (7.18) by $\tilde{U}^{As}$,$\tilde{U}^g$ and $\mathbf{c}^A$ respectively.

Both M_0 and M_1 are statistically based craniofacial models. Face landmarks and tissue depths are transformed in a face-specific way only while remaining in between statistically determined boundaries in order to guarantee facial plausibility. The transformation and plausibility is based on the knowledge of multiple reference faces within a database. The following two models, presented in the next section, use a single, instead of multiple, reference face, therefore throwing away a lot of covariance knowledge.

7.3.2 TPS based

Third model (M_2)

The third craniofacial model M_2 only uses the covariance of property values with other facial information parts in order to create a property normalized facial database. However, variances and covariances of face landmarks with tissue depths and complete textured facial surfaces are eliminated, resulting in the average normalized face or a specific normalized face of the database being the craniofacial model template with no knowledge of face-specific deformations and plausibility. Instead, a generic transformation model based on TPS is used as in chapter 5. We select the average normalized face $\bar{\tilde{F}}$ as model template such that the average normalized facial landmarks are the floating surface points $\{\mathbf{r}_j^f = \bar{\tilde{\mathbf{r}}}_j^{lm} | j = 1, \ldots, L\}$. The floating surface point transformation for the third model (M_2) is similar to (5.2) and (6.12):

$$\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}) = R\bar{\tilde{\mathbf{r}}}_j^{lm} + \mathbf{t} + \boldsymbol{\phi}(\bar{\tilde{\mathbf{r}}}_j^{lm})\Lambda \tag{7.22}$$

and the predefined distances, being the average normalized tissue depths, are not changed due to the lack of covariance knowledge:

$$d_j(\boldsymbol{\theta}) = \overline{\tilde{d}}_j \tag{7.23}$$

The complete set of parameters is now $\boldsymbol{\theta} = [\omega_x, \omega_y, \omega_z, t_x, t_y, t_z, \Lambda]$, while the regularization becomes:

$$\|\mathfrak{L}(\boldsymbol{\theta})\|^2 = \text{trace}(\Lambda^{\mathrm{T}}\Phi\Lambda) \text{ with } \Phi_{ij} = \|\overline{\tilde{\mathbf{r}}}_i^{lm} - \overline{\tilde{\mathbf{r}}}_j^{lm}\| \tag{7.24}$$

This is a smoothness regularization on the non-rigid deformation parameters Λ preventing the non-rigid subspace from behaving too arbitrarily. As already mentioned in chapter 5, the TPS kernel Φ is built based on the floating surface points containing information about their internal structural relationship. Using a TPS based generic transformation the prior knowledge or model-likelihood used during skull registration consists of a generic smoothness constraint, regularizing the generic deformation of the template. Alternatively, the craniofacial model compared to which the most plausible facial solution is searched for during registration is a single reference face. Finally, the textured facial surface estimate or reconstruction after model registration is computed according to:

$$\begin{aligned} F_{est}^s &= \{R\overline{\tilde{\mathbf{r}}}_i^s + \mathbf{t} + \boldsymbol{\phi}(\overline{\tilde{\mathbf{r}}}_i^s)\Lambda | i = 1, \ldots, N\} \\ F_{est}^g &= \{\overline{\tilde{g}}_i | i = 1, \ldots, N\} \end{aligned} \tag{7.25}$$

The registration result based only on the facial landmarks is generically and smoothly interpolated in order to obtain a final reconstruction surface result after skull registration. The texture values are equal to the averaged normalized texture values and are not changed according to the facial surface due the lack of proper covariance knowledge.

Fourth model (M_3)

The final craniofacial model M_3 is very similar to the previous one except that the craniofacial model template is now the average of the non-normalized database, not using any covariance information at all. The details for this model are the same as in (7.22),(7.24) and (7.25) except that the average property normalized template information $\bar{\tilde{F}}$ is replaced by the average non-normalized template information $\bar{F}$.

Note that these two TPS based models (M_2, M_3) are similar to more traditional craniofacial models found in literature. The first (M_2), uses a generic or specific model template adapted towards a given set of skull properties and the second (M_3) uses a non-adapted generic model template. Both combine the model template with a generic deformation model, such that prior knowledge is restricted to a single reference face combined with a smoothness constraint. Similar to the previous two statistical based craniofacial models these two models are perfectly applicable within the registration framework of section

7.2. The same gradient descent optimization is used due to the fact that the predefined distances are not equal to zero (making the ICP approach difficult to apply) and because the amount of floating surface points, influencing the amount TPS transformation parameters, is limited.

7.4 Unknown skull representations

The choice of skull representation defines the current distance $\mathfrak{D}(S^t, \mathbf{r})$ computation and the individual point similarity measures, therefore influencing the inlier-process component within the data-likelihood term of (7.2). The only major prerequisite for a skull representation to be appropriate is its compatibility with the craniofacial model. In our case this means that one has to be able to calculate the current distances within the 52 face landmarks of the model template. In this section we give details of two such model compatible skull representations usable within the registration framework. The first and primary skull representation that we propose is an implicit skull representation. The second is a more traditional craniometric skull representation based on manually indicated craniometric or anatomical skull landmarks.

7.4.1 Implicit

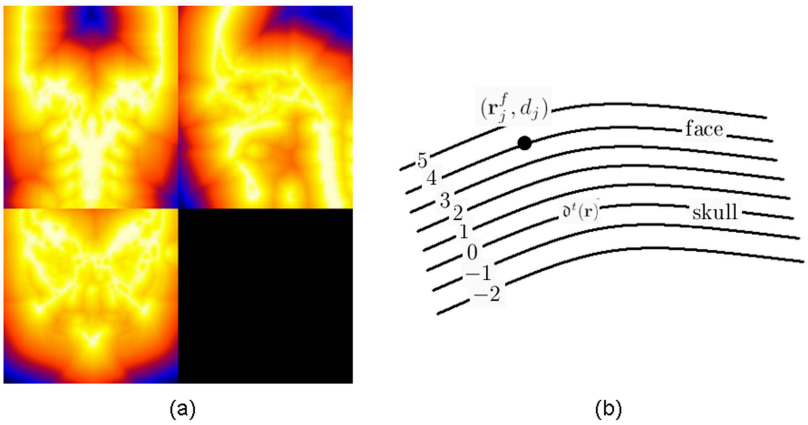


Figure 7.1: Illustration of the implicit skull representation $S^t = \mathfrak{d}^t(\mathbf{r})$. (a) Three slices trough the VIS of the skull visualized with a color map coding distances to the skull surface. (b) The relationship between a face landmark and the VIS.

The first and primary choice is an implicit skull representation, depicted in figure 7.1, leading to a fully automatic reconstruction because no manual interaction is required. The implicit or signed Distance Transform representation of the skull, being the target surface S^t, is obtained using the VIS machinery

of section 3.2.2 $S^t = \mathfrak{d}^t(\mathbf{r})$ and encodes for every 3D point an approximate shortest euclidian distance perpendicular to the skull surface, zero on the surface, positive outside and negative inside. The implicit surface based individual point similarities used during registration are based on the comparison of the signed distance transform values, being the current distance computations $\mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) = \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))$, with the ultrasound tissue depths, being the predefined distances $d_j(\boldsymbol{\theta})$, in the 52 facial landmarks, being the floating surface points $\mathbf{r}_j^f$:

$$\mathfrak{s}_j^*(\boldsymbol{\theta}) = \mathfrak{s}_j(\boldsymbol{\theta}) + \epsilon = \mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) - d_j(\boldsymbol{\theta}) + \epsilon \tag{7.26}$$

A face landmark $\mathbf{r}_j^f$ must be positioned such that the VIS value in the face landmark $\mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta}))$ is equal to the thickness value $d_j(\boldsymbol{\theta})$ of the face landmark. However, distances based on a VIS and distances measured with ultrasound are not exactly equal. The inequality between them is modeled as additive gaussian distributed noise ϵ with standard deviation σ. Therefore the inlier distribution for the implicit skull representation is:

$$p(S^t|\Theta) = \prod_{j=1}^{N^f} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(\mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) - d_j(\boldsymbol{\theta})\right)^2\right) \tag{7.27}$$

with N^f the number of floating surface points equal to the number of facial landmarks $N^f = L$. Furthermore, besides noise handling, outlier handling is needed as well for two reasons. Firstly, for landmarks situated at the forehead area, for example, both ultrasound and VIS values are very similar, but for landmarks situated in the cheek area both values are very dissimilar. Secondly, a CT scan of a skull with amalgam teeth filling will generate a disrupted VIS in the area of the teeth. Using (7.27) combined with outlier handling, the first term in the CFR objective function (7.12) during the M-step, becomes:

$$\begin{aligned}\mathfrak{S}(\Theta) = \quad & \textstyle\sum_{j=1}^{N^f} b_j(\widehat{\Theta}^{(m)}) \left[\frac{1}{2\sigma^2}\left(\mathfrak{d}^t(\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) - d_j(\boldsymbol{\theta})\right)^2 + \log\sqrt{2\pi}\sigma\right] - \\ & \textstyle\sum_{j=1}^{N^f}(1 - b_j(\widehat{\Theta}^{(m)}))\log\delta\end{aligned} \tag{7.28}$$

with $\Theta = \{\boldsymbol{\theta}, \sigma\}$. $b_j(\widehat{\Theta}^{(m)})$ is obtained during the E-step using (7.27) in (7.11). Details concerning the second term $\nu\|\mathfrak{L}(\theta)\|^2$, the floating surface point transformation $\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})$ and the predefined distance alternation $d_j(\boldsymbol{\theta})$ are dependent on the choice of craniofacial model and were given previously in section 7.3.

7.4.2 Craniometric

The craniometric skull representation $S^t = \{(\mathbf{s}_j^{lm}, \mathbf{n}_j^{lm}) | j = 1, \ldots, L\}$ used for CFR is based on L manually indicated craniometric skull landmarks $\mathbf{s}_j^{lm}$ (anatomically corresponding with the L face landmarks in the model) and

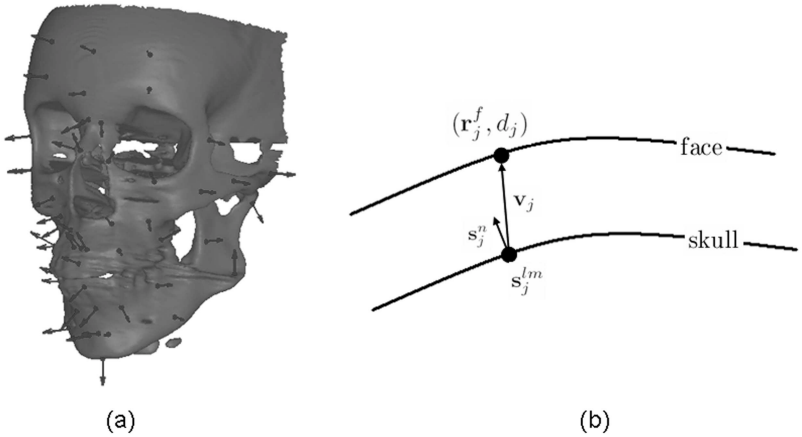


Figure 7.2: Illustration of the craniometric skull representation. (a) Skull surface with indicated craniometric skull landmarks and normal information. (b) The relationship between a face landmark and a corresponding skull landmark

surface normal information $\mathbf{n}_j^{lm}$, illustrated in figure 7.2. Current distances $\mathfrak{D}(S^t, \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})) = \|\mathbf{s}_j^{lm} - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\|$ are then obtained by calculating the distance between face landmarks $\mathbf{r}_j^f$ and corresponding skull landmarks (which are known prior to registration). The craniometric individual point similarities used during registration are based on the comparison of the distance $\|\mathbf{s}_j^{lm} - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\|$ with the ultrasound tissue depths $d_j(\boldsymbol{\theta})$, in the 52 facial landmarks $\mathbf{r}_j^f$:

$$\mathfrak{s}_j^{*d}(\boldsymbol{\theta}) = \mathfrak{s}_j^d(\boldsymbol{\theta}) + \epsilon^d = \|\mathbf{s}_j^{lm} - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\| - d_j(\boldsymbol{\theta}) + \epsilon^d \tag{7.29}$$

A face landmark $\mathbf{r}_j^f$ must be positioned such that the distance to the corresponding skull landmark is equal to the thickness value of the face landmark. However, as mentioned in section 3.3.2, having predefined distances not equal to zero, the similarity measures in 7.29 are ill-defined. Spherical possibilities for the repositioning of the face landmarks are created, resulting in a disrupting of the target skull surface geometry or shape. To solve this problem, additional target surface information and prior knowledge besides corresponding point positions is to be incorporated within the similarity measure favoring certain positions along the spheres of possibilities. Knowing that ultrasound tissue depth measurements were taken perpendicular to the skull surface, guaranteeing the best ultrasound signal, we can use the skull surface normal information $\mathbf{n}_j^{lm}$ in the skull landmarks to favor spherical possibilities along the positive direction of the normal. Doing so, the individual point similarity measure in

(7.29) is augmented with an additional individual point similarity measure:

$$\mathfrak{s}_j^{*a}(\boldsymbol{\theta}) = \mathfrak{s}_j^a(\boldsymbol{\theta}) + \epsilon^a = 1 - \frac{\mathbf{v}_j(\boldsymbol{\theta}) \cdot \mathbf{n}_j^{lm}}{\|\mathbf{v}_j(\boldsymbol{\theta})\| \|\mathbf{n}_j^{lm}\|} + \epsilon^a \tag{7.30}$$

with $\mathbf{v}_j(\boldsymbol{\theta}) = \mathbf{s}_j^{lm} - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})$ the measuring direction. Using (7.29) combined with (7.30), a face landmark must be positioned at a distance to the corresponding skull landmark equal to the thickness value in the face landmark, while the angle between the measuring direction, which is the direction of the line trough the skull landmark and the face landmark, and the normal direction in the skull landmark must be equal to zero. Doing so, the craniometric based similarity criterium expressing the goodness of fit between this skull representation and a set of face landmarks is based on the procedure for measuring the tissue depths in the database with ultrasound measuring technology. Note that the additional similarity measure is not needed using the implicit skull representation because mapping possibilities are restricted to lay on iso-distance surfaces retaining the target surface geometry at different distances from the zero-distance surface.

However, the indicated craniometric skull landmarks are not error free. Besides the fact that the virtual copy of the skull is only an approximation of the real skull specimen, additional errors are introduced during skull landmark indication. These errors have an effect on the first part of the similarity criterium (7.29) and are modeled as additive gaussian distributed noise ϵ^d with standard deviation σ_d. Furthermore, the normals in the skull landmarks are only approximations of the real surface normal. These errors have an influence on the second part of the similarity criterium (7.30) and are modeled as additive gaussian distributed noise ϵ^a with standard deviation σ_a. Suppose that the distance error ϵ^d and the angle error ϵ^a are independent then the inlier distribution for the craniometric skull representation can be defined based on the product of two normal distributions:

$$p(S^t|\Theta) = \prod_{j=1}^{N^f} p(\mathfrak{s}_j^d|\Theta) p(\mathfrak{s}_j^a|\Theta) \tag{7.31}$$

with

$$p(\mathfrak{s}_j^d|\Theta) = \frac{1}{\sqrt{2\pi}\sigma_d} \exp\left(-\frac{1}{2\sigma_d^2}\left(\|\mathbf{s}_j^{lm} - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\| - d_j(\boldsymbol{\theta})\right)^2\right) \tag{7.32}$$

and

$$p(\mathfrak{s}_j^a|\Theta) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left(-\frac{1}{2\sigma_a^2}\left(1 - \frac{\mathbf{v}_j(\boldsymbol{\theta}) \cdot \mathbf{n}_j^{lm}}{\|\mathbf{v}_j(\boldsymbol{\theta})\| \|\mathbf{n}_j^{lm}\|}\right)^2\right) \tag{7.33}$$

N^f is the amount of floating surface points equal to the number of facial landmarks $N^f = L$. Besides noise handling, outlier handling is needed as

well due to the fact that certain skull landmark indications and normal computations can contain gross errors due to the manual intervention which is dependent on the expertise of the anthropologist performing the reconstruction. Having two individual point similarities the function defined in (7.6) to obtain an equivalent M-estimator becomes the sum of two similar functions $h(x, \alpha(z), \beta(z)) = h^d(x^d, \alpha(z), \beta(z)) + h^a(x^a, \alpha(z), \beta(z))$ using appropriately σ_d and σ_a in (7.7) and with x^d and x^a equal to $\mathfrak{s}_j^d(\boldsymbol{\theta})$ and $\mathfrak{s}_j^a(\boldsymbol{\theta})$, respectively. Using (7.31),(7.32) and (7.33) combined with outlier handling, the first term in the CFR objective function (7.12) during the M-step, being the conditional expectation of the complete negative log data-likelihood becomes:

$$\begin{aligned}\mathfrak{S}(\Theta) = \quad & \textstyle\sum_{j=1}^{N^f} b_j(\widehat{\Theta}^{(m)}) \left[\frac{1}{2\sigma_d^2} \left(\|\mathbf{s}_j^{lm} - \mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})\| - d_j(\boldsymbol{\theta}) \right)^2 + \log \sqrt{2\pi}\sigma_d \right] + \\ & \textstyle\sum_{j=1}^{N^f} b_j(\widehat{\Theta}^{(m)}) \left[\frac{1}{2\sigma_a^2} \left(1 - \frac{\mathbf{v}_j(\theta) \cdot \mathbf{n}_j^{lm}}{\|\mathbf{v}_j(\theta)\| \|\mathbf{n}_j^{lm}\|} \right)^2 + \log \sqrt{2\pi}\sigma_a \right] - \\ & \textstyle\sum_{j=1}^{N^f} (1 - b_j(\widehat{\Theta}^{(m)})) \log \delta \end{aligned} \tag{7.34}$$

with $\Theta = \{\boldsymbol{\theta}, \sigma_d, \sigma_a\}$. $b_j(\widehat{\Theta}^{(m)})$ is obtained during the E-step using (7.31) in (7.11). (7.34) seems a bit overwhelming, but is perfectly differentiable for optimization purposes. Similar to the implicit skull representation, details concerning the second term $\nu\|\mathfrak{L}(\theta)\|^2$ being the negative log model-likelihood in (7.12), the floating surface point transformation $\mathfrak{T}(\mathbf{r}_j^f, \boldsymbol{\theta})$ and the predefined distance alternation $d_j(\boldsymbol{\theta})$ are dependent on the choice of craniofacial model and were given previously in section 7.3.

7.5 Results and validation

In this section results and a validation, concerning the CFR framework, the craniofacial models and the skull representations, are given. Therefore a validation framework is proposed to substantiate on a scientific basis the added value of reconstruction methods during crime-scene investigations. Issues we want to analyze concern the added value of a statistical model for CFR, the use of an implicit versus a traditional craniometric unknown skull representation and the performance of noise and outlier robustness in the proposed CFR framework.

A proper craniofacial reconstruction validation requires skull information, according ground truth facial knowledge and a facial comparison methodology. Furthermore, in order to be representative, validation should not be limited to a single case. Therefore we constructed and applied two different validation databases. The first database is a clinical patient database and is used to simulate realistic reconstruction scenarios. The second database, being the database from which the model is build, is used for simulating controlled noisy

and outlier skull data points in a craniometric skull representation. Following a cross-validation approach, each facial sample can be removed, in turn, from the databases and used as a test case. The resulting reconstructed facial skin surface can then be compared quantitatively and qualitatively, using the facial comparison methodology, with the skin surface of the test case representing the ground truth. Doing so for all the faces within a database, overall reconstruction performances can be obtained.

We start with the explanation of the facial comparison methodology. Then the results based on the patient database are given and discussed after which we do the same for the model database. We end with two final teasers, a victim identification case and an anthropological Neanderthal reconstruction.

7.5.1 Facial comparison

The purpose of the facial comparison is to give a quantitative and qualitative reconstruction evaluation. The quantitative evaluation is based on local surface differences and expresses the reconstruction performance in terms of accuracy. Furthermore a spatial map of the reconstruction accuracies of each facial region can be provided. However, the final goal of craniofacial reconstruction is not reconstruction accuracy, but rather recognition or identification success. Therefore, a qualitative measure, based on a statistical surface model, is computed and expresses the reconstruction performance in terms of recognition.

Quantitative evaluation

Suppose we have a set of facial reconstructions $\{\mathbf{F}_l^R | l = 1, \ldots, R\}$ with according ground truth (GT) facial information $\{\mathbf{F}_l^{GT} | l = 1, \ldots, R\}$. The quantitative evaluation of a single reconstruction is based on the distances $d_{li} = \|\mathbf{r}_{li}^{sR} - \mathbf{r}_{li}^{sGT}\|$, being local surface errors, between corresponding points $\{(\mathbf{r}_{li}^{sR}, \mathbf{r}_{li}^{sGT}) | i = 1, \ldots, N\}$ on the l^{th} reconstruction surface and the l^{th} GT surface. Note that corresponding points are not equal to closest points, such that the local errors are not underestimated. These local surface errors are used to determine a global root mean squared error value ($rmse_l = \sqrt{\sum_{i=1}^{N} d_{li}^2 / N}$) for a single reconstruction. Doing this for all the reconstruction we obtain an averaged overall global rmse ($rmse = \sum_{l=1}^{R} rmse_l / R$). Additionally we also compute an overall local rmse ($rmse_i = \sqrt{\sum_{l=1}^{R} d_{li}^2 / R}$, which can be visualized using the shape of the average face for example.

Qualitative evaluation

In order to examine the performance in terms of 3D face identification, we construct a statistical facial surface model $M^s(\mathbf{g}) = \overline{\mathbf{F}^s} + U^s \mathbf{g}$ with $\mathbf{g} = \{g_j | j = 1, \ldots, D\}$ based on the facial surfaces $\{\mathbf{F}_k^s | k = 1, \ldots, K\}$ within the model database. Based on this facial surface model we have a set of parameters

$\{\mathbf{g}_k | k = 1, \ldots, K+1\}$ for every face in the database including the GT facial surface, where $\mathbf{g}_{K+1} = \mathbf{g}^{GT}$ is obtained according to:

$$\mathbf{g}_{GT} = (\mathbf{F}^{sGT} - \overline{\mathbf{F}^s})/U^s \tag{7.35}$$

Additionally, the model description $\mathbf{g}_R$ of a reconstruction is obtained as well similar to (7.35).Then we compare the reconstruction with every face in the database including the GT using a similarity measure or distance between faces in the face-space defined in [104] based on the cosine of the angle between scaled model descriptions $\widehat{\mathbf{g}} = \{g_j/\sigma_j | j = 1, \ldots, D\}$ according to the standard deviations σ_j being the square root of the eigenvalues belonging to the eigenvectors in U^s.

$$d_A(\breve{\mathbf{g}}_k, \breve{\mathbf{g}}_R) = 1 - \frac{\breve{\mathbf{g}}_k \cdot \breve{\mathbf{g}}_R}{\|\breve{\mathbf{g}}_k\| \|\breve{\mathbf{g}}_R\|} \tag{7.36}$$

The smaller $d_A(\breve{\mathbf{g}}_k, \breve{\mathbf{g}}_R)$ between the reconstruction and a face from the database the more alike the two are. Based on this similarity measure a rearranging of the database can occur starting with the most similar face and ending with the most dissimilar face. If the GT facial surface is situated on the n^{th} place in this rearranged database, then we say that the reconstruction recognition success is of rank n. Of course a rank 1 reconstruction is a perfect recognition. Because these tests are dependent on the batch size (the number of faces in the database) the rankings are expressed in percentage. A rank 1% is a reconstruction for which the GT is situated within the first percentage of the rearranged database. A typical visualization of recognition performance is the cumulative recognition curve. On the horizontal axes (X) the rank in percentage of the reconstructions is shown, while on the vertical axes (Y) the percentage of reconstructions having a rank smaller than or equal to the rank on the X axes is shown. A good recognition performance curve is one with a high Y% value for the 1% X value and a small X% value for the 100% Y value.

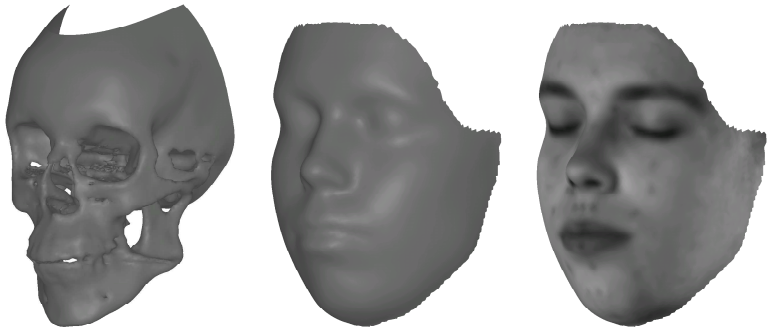
Having the material to evaluate a reconstruction both in terms of accuracy and recognition we can elaborate the tests and results based on the patient database and the model database. We start with the patient database in the next section.

7.5.2 Patient database

Material

A first set of validation results of the proposed craniofacial reconstruction method is based on a database of 12 clinical patient cases. For every case we acquired the complete skull surface using a clinical CT scanning protocol. The skin surface in standard upright position is acquired with the photogrammetric 3D scanner (ShapeCam, Eyetronics). Subsequently, the statistical model of chapter 6 is fitted to the skin surface according to section 6.4.1, generating a model description for every case as in figure 7.3. Each facial sample from the validation database is used as a test case. The resulting reconstructed facial

skin surfaces can then be compared quantitatively and qualitatively, with the model skin surface description of the test case representing the ground truth. Thanks to the model based GT registration, corresponding points are obtained to perform the quantitative facial comparison.



(a) (b) (c)

Figure 7.3: Ground truth illustration for a case from the patient database. (a) skull acquired with a CT scanner. (b) GT facial surface obtained by a model-based facial surface registration onto the 3D eyetronics acquisition. (c) similar as (b) but rendered with texture information.

For every case we register the four different craniofacial models (M_0, M_1 , M_2, M_3) using the EM craniofacial reconstruction framework over a range of kappa values $\kappa = [1.5; 3.5]$ in steps of 0.2 based on two craniometric (CM) point based skull representations and the implicit (VIS) representation of the skull. The first craniometric representation (CM_1) is an indication of 52 skull landmarks without knowledge of the GT face belonging to the skull. The second (CM_2) is a better indication with knowledge of the GT. For both craniometric representations the normals on the skull surface are determined still incorporating errors for reasons explained earlier, so CM_2 is considered a good but not perfect indication. The user-specific amount of regularization is set equal to one ($\nu = 1$) for the statistical based craniofacial models and $\nu = N^f$ for the TPS based models. A total amount of 4(models) x 3(representations) x 11(κ values) reconstructions is made for every case.

Global rmse and kappa influence

A quantitative evaluation for every reconstruction is determined such that we obtain an overall global rmse (averaged over the 12 cases) for a certain reconstruction setup (choice of model, representation, kappa value), which is depicted in figure 7.4. From these results, we see that the performance of the TPS based models M_2 and M_3 is inferior compared to the statistical based models M_0 and M_1. Concerning the kappa value influence a few things can be

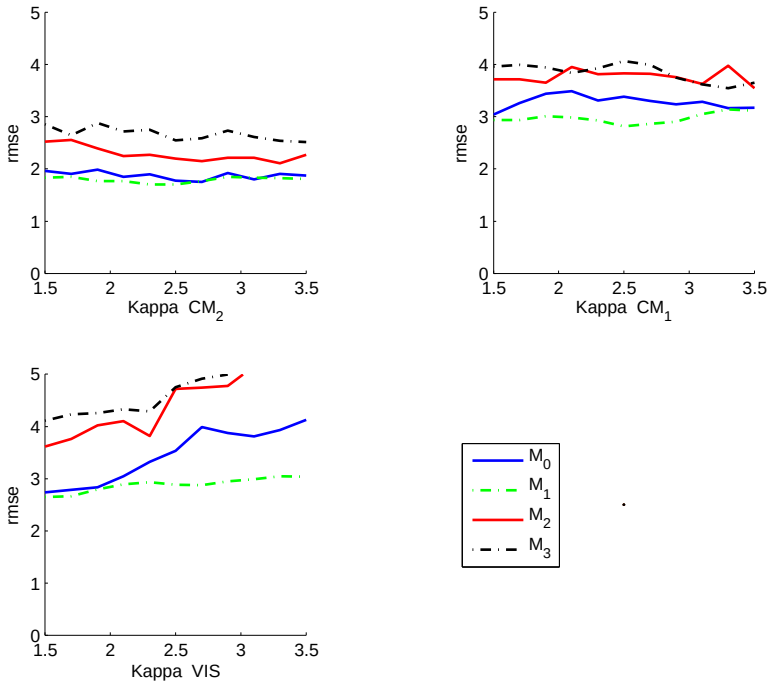


Figure 7.4: Vertical axes: overal global RMSE over 12 cases for a reconstruction setup. Horizontal axes: kappa value. Top Left: CM_2 skull representation. Top Right: CM_1 skull representation Bottom Left: VIS skull representation. solid Blue curves: M_0, dotted Green curves: M_1 , solid Red curves: M_3, dotted Black curves: M_2

observed. For the CM_2 representation the results are quite constant over the whole kappa range, indicating that no severe outliers are present in this skull representation, which was predictable knowing that the CM_2 representation is based on a good craniometric landmark indication. The choice of kappa value for this representation is of low importance. The results in terms of kappa value based on the CM_1 seem to be similar, but the curve stabilities are less pronounced, indicating the presence of more outliers compared to the CM_2 representation. However it is hard to make any conclusions for this representation in terms of kappa value, because the quality of the manual indicated landmarks over the 12 cases is very variable and not equal for all cases. The VIS representation on the other hand is very similar over the 12 cases in terms of quality, so that a proper kappa value analysis is possible. The curves in figure 7.4 (Bottom Left) show an increase in rmse with increasing kappa values, suggesting lower kappa values for better reconstructions based on this implicit representation. The increase in rmse is due to the fact that we compare VIS values with ultrasound thickness values, which are not measured entirely the

same way. For landmarks situated around the forehead for example both values are very similar, but for landmarks situated in the cheek area both values are very dissimilar. The implicit skull representation used in combination with ultrasound thicknesses is characterized by a severe amount of outliers. Increasing the kappa value comes down to believing or trusting the VIS/ultrasound value relationship in more landmarks including outliers, influencing the CFR registration wrongfully. Another observation is that the model M_1 is less influenced by the choice of kappa value compared with model M_0. This is due to the fact that thickness values and facial geometry change independently using M_1, while a co-varying change is used within M_0 during registration. An outlier VIS/ultrasound value relationship influences both thickness and surface estimates within M_0, while the influence is mostly limited to the thickness estimate within M_1. Combined thickness and facial surface plausibility decreases more rapidly than independent thickness and facial surface plausibility with increasing believe in the VIS/ultround value relationship.

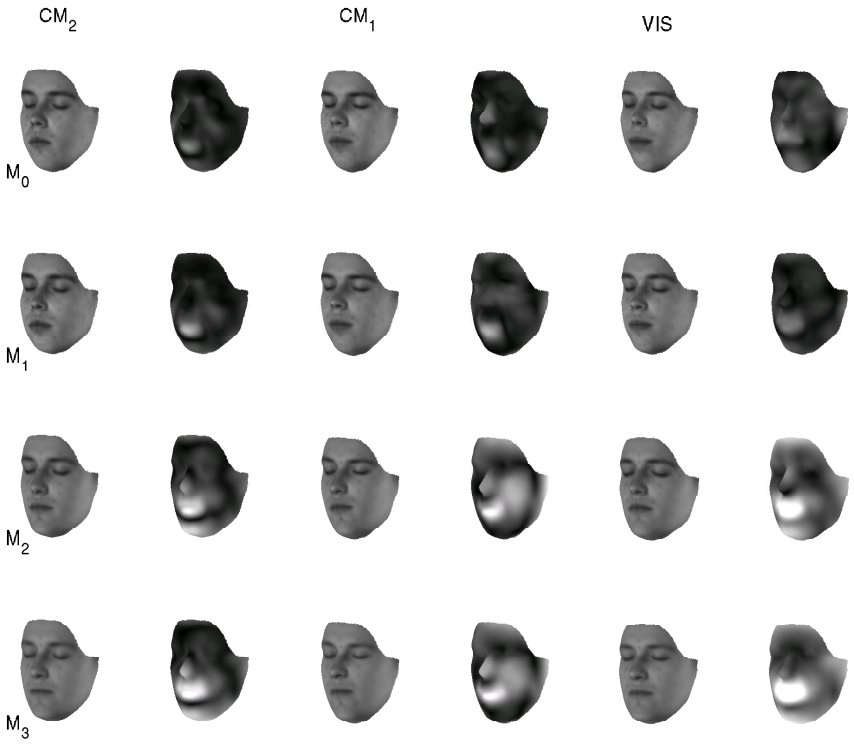


Figure 7.5: Local surface reconstruction results for the case depicted in figure 7.3

Local rmse accuracy

For further analysis we select the reconstruction with the best kappa value for every combination of model and skull representation per case. The best κ value is the one having the lowest rmse of all reconstructions made over the κ range, such that a single reconstruction becomes the representative for a certain case reconstruction setup consisting of the choice of model and skull representation. The local surface differences of these representative reconstructions can be visualized using the geometry of the reconstructed face in combination with a color-map. The results for the single case in figure 7.3 are depicted in figure 7.5 by means of a color-map ranging from 0 mm(black) to 7 mm (white). The first two columns show the results obtained based on the CM_2 skull representation made with the four different models ($M_0 - M_3$). The following two columns show the results obtained based on the CM_1 craniometric skull for each of the four different models ($M_0 - M_3$). The final two columns show the results obtained based on the implicit VIS skull representation made with the four different models ($M_0 - M_3$). For 12 cases we obtain overall local rmse surface differences, shown in figure 7.6 using the averaged facial geometry of the database. The ordering of the figure is the same as figure 7.5, except that the rows and columns are switched. Figures 7.6 and 7.5 use the same color-map range. Note that a small difference in overal global rmse as (figure 7.4) can incorporate more pronounced differences in overall local surface rmse errors (figure 7.6). This is important while analyzing overall global rmse errors. It is observed that the nose and the mouth region are the most difficult to reconstruct for all models and skull representations, which is due to the limited underlying skull information in those regions.

Recognition

Comparing two surfaces (reconstruction vs GT) based on local surface differences is interesting for evaluating the reconstruction performance in terms of accuracy and provides a spatial map of the difficulty of each facial region to be reconstructed. However, the final goal of cranio-facial reconstruction is not reconstruction accuracy per se, but rather recognition or identification success. Using the qualitative facial comparison, the ranking results in percentage of the representative reconstructions for every case are shown in figure 7.7. A more typical visualization of recognition performance is depicted in figure 7.8 using cumulative recognition curves.

Model and skull representation comparisons

Observing figures 7.5, 7.6, 7.7 and 7.8 a few conclusions can be drawn concerning the craniofacial models and skull representations. A global observation is the fact that accuracy is related with recognition and vice versa.

CFR models: Starting with M_3 we see that this model has the worst performance both in terms of identification and accuracy. Especially the mouth, chin and nose regions are poorly reconstructed (figure 7.6). If we compare M_3

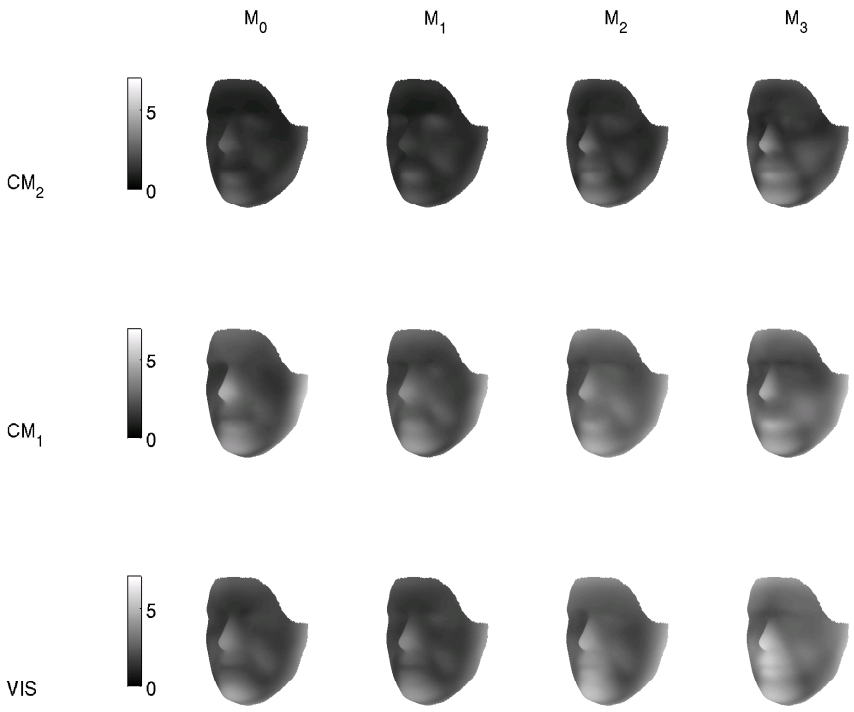


Figure 7.6: Overall local rmse surface reconstruction results over 12 cases, depicted on the geometry of the average face from the database

with M_2 in all the validation figures we clearly see that the incorporation of attribute or property covariance gives a positive contribution to the quality of the reconstructions both in terms of accuracy and in terms of recognition. It is clearly necessary to obey the property values, by manipulating the craniofacial model to improve the results. Comparing M_2 (statistical-based) to M_1 (TPS-based) it is observed that a major positive contribution is obtained by using face-specific deformations, extracted from facial surface and landmark covariance information and encoded into the eigenvectors. Incorporating statistical knowledge of multiple reference faces is clearly better than using only one reference face in combination with a generic TPS deformation. Furthermore, we believe, however not proven by the current tests, that reconstructions made based on a combination (weighted average e.g.) of multiple reconstructions starting from several reference faces in combination with generic deformations incorporate the danger that the final result is corrupted by possible inferior single reconstructions as the ones based on M_3 or M_2. For this reason, we prefer to make a single reconstruction based on the statistical knowledge of multiple faces during registration instead of combining multiple reconstructions after registrations. Finally, comparing M_1 with M_0, only a slight difference

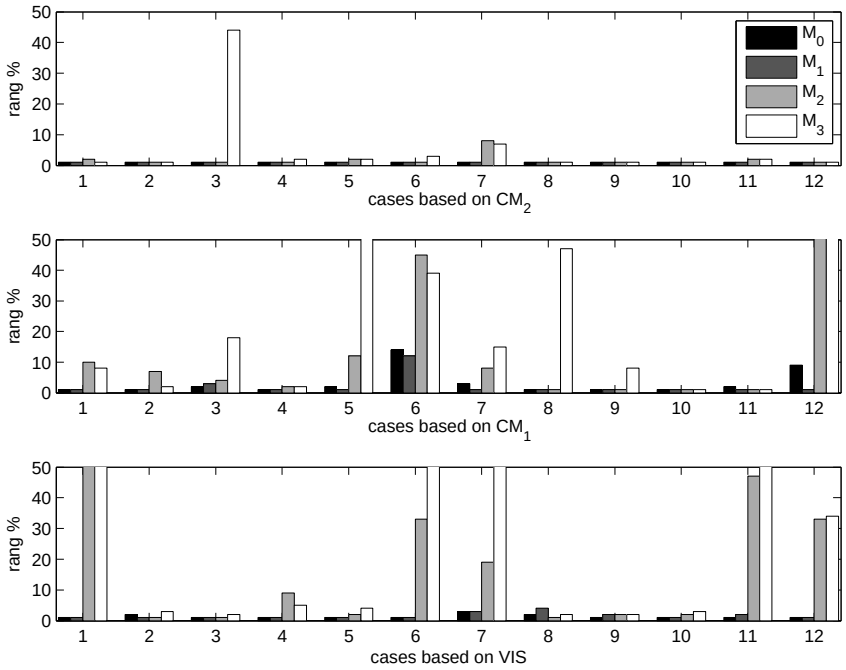


Figure 7.7: % rankings for the 12 cases of the best representative reconstructions for the CM_2, CM_1, VIS skull representations and the 4 different models (M_0, M_1, M_2, M_3)

is observed, sometimes in favor of M_1 and sometimes in favor of M_0. Based on these tests, it is not really clear whether and/or when the incorporation of thickness covariance improves or reduces the reconstruction quality. More controlled tests comparing M_1 and M_0 are given further on. Concerning the different craniofacial models $(M_0 - M_3)$ we can conclude for now, that the incorporation of property and facial surface/landmark covariance improves the reconstruction quality both in terms of accuracy and in terms of recognition performance.

Unknown skull representations: Concentrating on the statistical models M_1 and M_0, we can make a few conclusions about the different skull representations as well. Both M_1 and M_0 generate excellent reconstructions based on the CM_2 representation. The cumulative recognition curves in figure 7.7 (top) are even flawless. However, the quality severely drops when using the CM_1 representation. This indicates that a reconstruction result based on a craniometric skull representation is very dependent on the quality of the landmark indication, which is related to the expertise of the anthropologist making the reconstruction. The results based on the implicit VIS representation are situated in between the results based on the CM_1 and CM_2 representations resulting into the following skull representation ordering: $CM_2 > VIS > CM_1$.

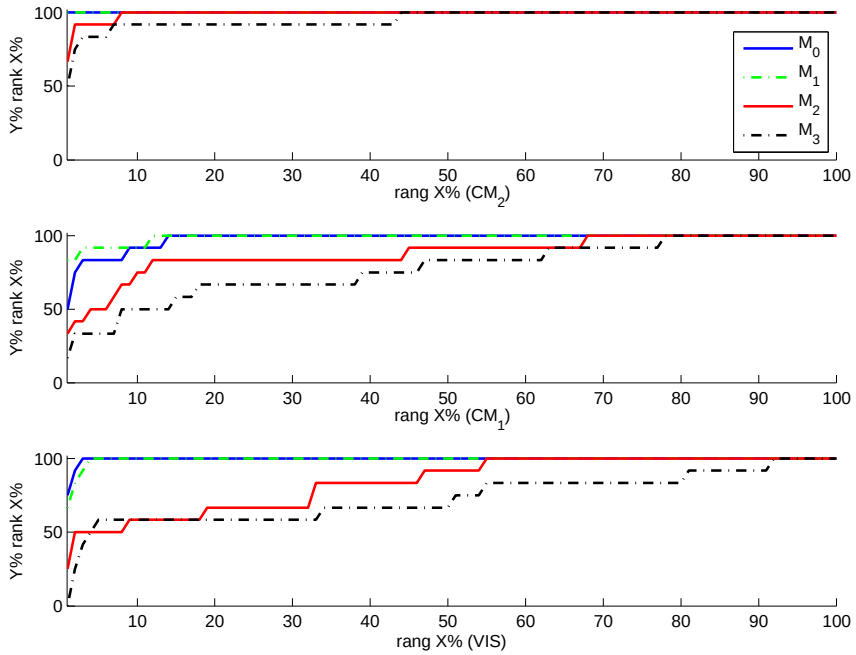


Figure 7.8: Cumulative recognition curves of the best representative reconstructions for the CM_2, CM_1, VIS skull representations and the 4 different models (M_0, M_1, M_2, M_3)

Observing figure 7.6 in terms of accuracy, we see that the spatial local error map based on the VIS representation for the models M_1 and M_0 is slightly better than the error maps for the CM_1 representation. In terms of identification, we observe in figure 7.7, that errors in recognition, when present, are less severe with the VIS representation resulting in a better cumulative recognition curve in figure 7.8. Keeping in mind that the CM_1 representation is based on a more realistic landmark indication in practice compared to CM_2 and that a completely automatic reconstruction, excluding human subjectivity, can be generated based on the VIS representation, we conclude that the VIS representation is a good skull representation to use in practice despite the weak comparison of ultrasound thicknesses with VIS values, resulting in a low kappa value choice. The reconstruction result based on the implicit representation, which is obtained completely automatically, can be used to generate an initial reconstruction. Subsequently, skull landmarks with some initial facial knowledge can then be indicated to refine the result if preferred.

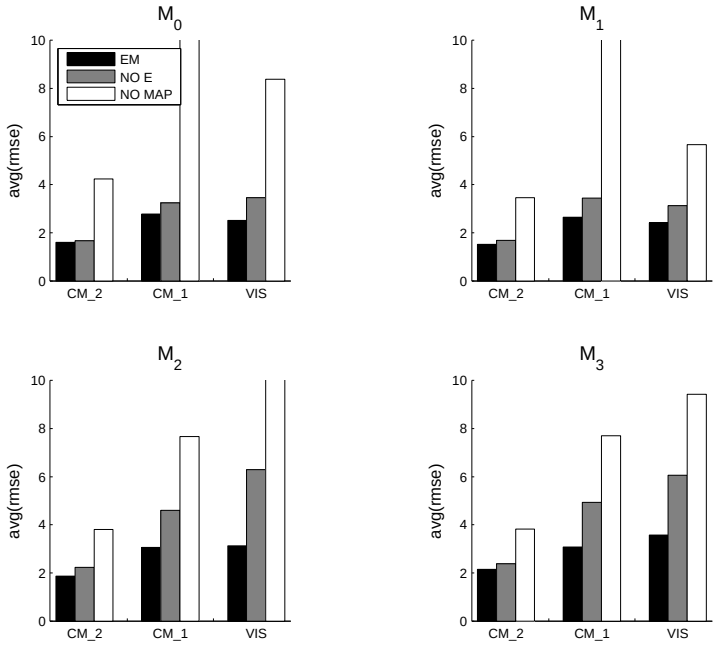


Figure 7.9: Overall global rmse results over the 12 cases, with the EM skull registration framework, Without the E-step and without the MAP formulation for every craniofacial model (M_0, M_1, M_2, M_3) and skull representations (CM_2, CM_1, VIS)

Robustness performance

To validate the contribution of the robust EM CFR framework, two more reconstructions per setup (choice of case, skull representation and craniofacial model) are made. The first eliminates the robustness against outliers by eliminating the E-step during registration. The latent variables z_j are set equal to one at initialization and are not updated during registration. The second eliminates the noise estimation σ of the skull representation by not using the MAP formulation of the registration problem. The results are shown in figure 7.9 using the overall global rmse over the 12 cases. For every model and skull representation the elimination of the E-step clearly decreases the reconstruction quality. This decrease in quality is less severe using the CM_2 skull representation compared to the other skull representations, because the amount and severity of outliers is lower within the CM_2 representation. Eliminating the estimation of skull representation noise during registration further decreases the reconstruction quality severely, even when using the outlier free but noisy CM_2 representation. Based on these comparisons we can conclude that it is important to deal with the noise in the skull representations and to deal with the presence of outliers compared to the noise level during registration. Fur-

thermore we observe that the proposed EM skull registration framework is a valuable tool to cope with such noise and outliers.

While these tests simulated realistic reconstruction scenarios, it is quite difficult to control the quality of the skull representation in terms of noise and outliers. Therefore we constructed additional tests based on the model database in the next section.

7.5.3 Model database

Material

A second set of validations is obtained using the facial database to build up the craniofacial models in order to have a better control of the noise and outliers present in the skull representation. Each facial sample is removed, in turn, from the model database and used as a test case resulting in 393 test cases. The remaining facial entries are used to create the different craniofacial models. Because of the lack of complete skull surface information in the model database, the implicit skull representation cannot be used. Instead we use a better controlled, compared to the patient database, craniometric skull representation. Starting from the face landmarks on the facial surface the thicknesses of the person are set out along the inward normal direction of the facial surface creating a set of skull landmarks and skull normals (being the facial normals). Using these skull landmarks and normals an noiseless relationship according to (7.29) combined with (7.30) exists with the face landmarks.

Noise robustness

To examine the influence of noise or small errors in the skull representation we gradually add gaussian distributed noise to the 3D positions of the skull landmarks. A skull landmark is displaced using a uniformly distributed random direction in combination with a gaussian distributed step-size in the random direction, resulting into spherical displacements of certain magnitudes. The amount of noise or the noise standard deviations are 0, 1, 2, 3 and 4 mm, resulting in five CM skull representations per face in the database. Every model (M_0, M_1, M_2, M_3) is used to make reconstructions over a kappa range $\kappa = [1.5; 3.5]$ in steps of 0.2 based on the five CM skull representations with different levels of inserted noise. Note that gaussian distributed noise can generate errors much bigger than the chosen standard deviation. In other words the maximum errors are not limited to 4 mm.

Figure 7.10 depicts the overall global rmse averaged over 393 cases as a function of kappa similar to figure 7.4. Based on these plots we select a single kappa value for every craniofacial model having the lowest overall global rmse, resulting in a representative reconstruction for every model and noisy skull representation. Doing so, we can construct figures 7.11 and 7.12 similar to figures 7.8 and 7.9 to compare the different craniofacial models and to validate the contribution of the EM skull registration framework in the case of noise present in the skull representation. Increasing the noise level or reducing

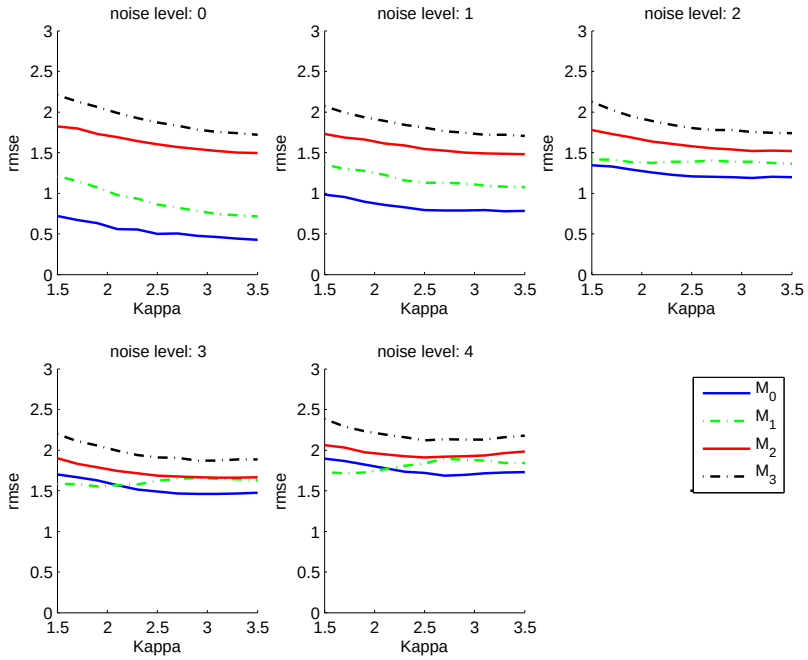


Figure 7.10: The overall global rmse results for five different noise levels in a CM skull representation in function of kappa. Vertical axes: overall global rmse over 393 cases for every model. Horizontal axes: kappa value.

the skull landmark quality clearly reduces the reconstruction qualities both in terms of accuracy and recognition, which was already observed comparing the CM_1 with the CM_2 skull representations in the clinical patient database. Furthermore, the conclusions made previously concerning the different craniofacial models based on the clinical patient database are again confirmed when observing figures 7.10, 7.11 and 7.12. However, a more detailed kappa analysis and model comparison can be given. A higher kappa value is preferred when the amount of noise is limited. Every skull landmark is quite trustworthy and gives a positive contributing to the noise level and model parameter estimation. No real outliers are present or were inserted. Increasing the noise level decreases slightly the optimal choice of kappa value. Still no severe outliers are inserted as such, but the chance of having skull landmarks of different quality is higher compared to low noise levels and as a result some skull landmarks are treated as outliers compared to the other skull landmarks. Trusting the complete set of skull landmarks when the noise level is high is not advisable, so a lower value of kappa is preferred. The statistical models M_0 and M_1 clearly increase the reconstruction quality when the amount of noise is limited compared to M_2 and M_3. The statistical interpolation (models M_0 and M_1) of the complete facial surface using co-varying surface information based on

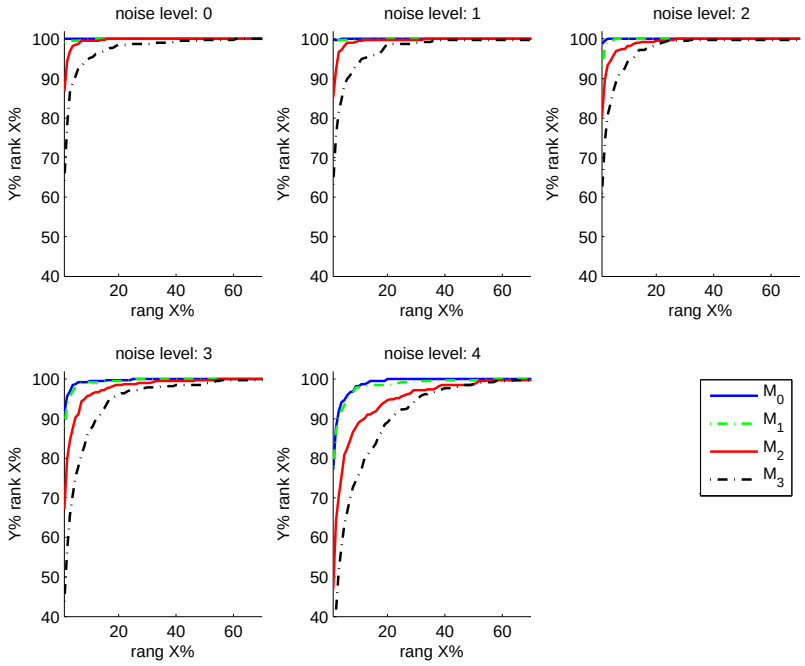


Figure 7.11: Cumulative recognition curves for 393 cases using the four different craniofacial models based on five different noise levels in the skull representation.

a set of face landmarks derived from quite noiseless skull landmarks is better than generic TPS based interpolation (models M_2 and M_3). Despite a smaller difference, the performance of M_0 and M_1 is still better with increasing noise. The difference between M_0 and M_1 is more pronounced compared to the clinical patient database validation tests. When the link between the skull representation and the model through the use of a similarity criterium is strong or not corrupted by a high amount of noise then the incorporation of thickness covariance (M_0) improves the reconstruction quality. Up to a noise level of 2mm the results based on M_0 are better both in terms of reconstruction accuracy (figures 7.10 and 7.12) and reconstruction identification success (figure 7.11). When the amount of noise increases we observe that the performance of M_0 and M_1 become comparable (figure 7.10). However, the performance of M_0 is less dependent on the choice of kappa value. A global Craniofacial transformation model regularization is needed to cope with very noisy skull representations and the flexibility of M_0 is more restricted than M_1 trough the combined thickness and facial surface plausibility. M_1 is allowed to interpolate the noisy skull landmarks more than M_0, which is not preferred. Lowering the kappa value is equal to mistrusting the noisy skull landmarks and down weighting the contribution of the most corrupted landmarks for the parame-

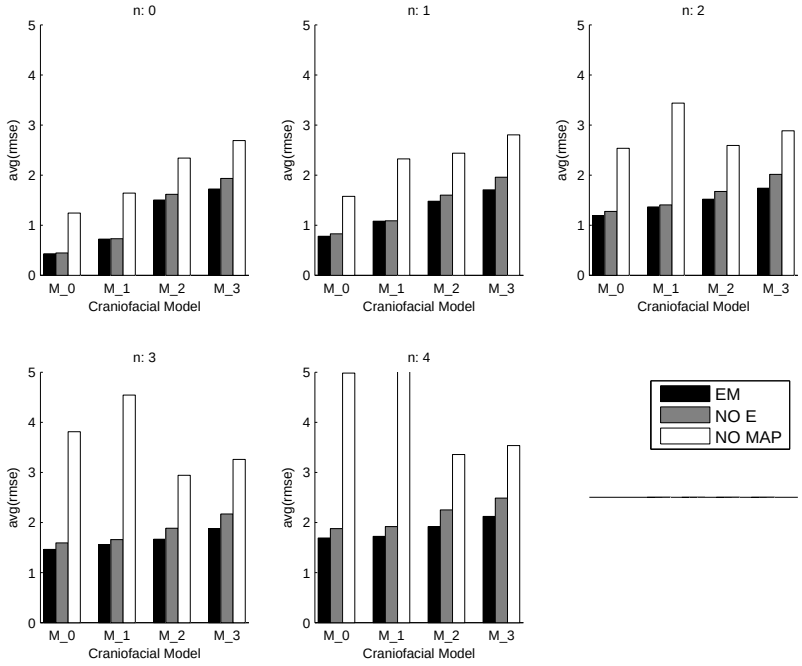


Figure 7.12: Overall global rmse over 393 cases of the representative reconstructions using the EM skull registration framework and the best kappa value, Without the E-step or outlier robustness and without the MAP formulation or noise robustness for every craniofacial model (M_0, M_1, M_2, M_3) and five CM skull representations with different noise levels

ter estimation. In that sense the choice of kappa value is an additional kind of regularization, but more individual skull landmark based, helping both M_0 and M_1 but mostly M_1 to improve the skull landmark approximation. Further increasing the noise level as in the CM_1 skull representation for the clinical patient database, can result in a slight improved performance of M_1 compared to M_0. Finally, observing figure 7.12 we clearly see an improvement in terms of noise robustness using the statistical MAP formulation (NO E) of the registration problem. Incorporating robust statistical estimation (EM) for outlier detection mainly improves the robustness within higher noise levels, but the contribution is limited because no severe outliers were incorporated into the simulated noisy skull representations. Using no noise robust registration (NO MAP) has a huge effect on the statistical models M_0 and M_1. Eliminating the statistical model regularization during registration can lead to facial instances outside the determined statistical boundaries, which have a very low facial plausibility shown in figure 7.13.

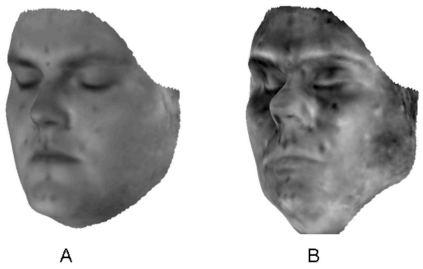


Figure 7.13: (a) A facial example generated based on noisy skull landmarks with the proposed EM framework and a statistical craniofacial model. (b) A facial example generated outside the statistical boundaries with very low facial probability based on the same noisy skull landmarks and statistical craniofacial model without the proposed EM framework for model registration

Outlier robustness

To examine the influence of outliers or gross errors in the skull representation we gradually insert an amount of outliers into a gaussian distributed noisy skull representation with standard deviation 1mm. An outlier skull landmark is selected randomly from the 52 skull landmarks and is displaced using a uniformly distributed random direction in combination with a fixed step-size of 6mm in the random direction. The choice of 6mm is made to ensure severe outliers compared to the noise level of 1mm. The amount of outliers, which are inserted are 0%, 5%, 10%, 20% and 40% of the 52 skull landmarks, resulting in five CM noisy skull representations with different amounts of outliers per face in the database. Every model (M_0, M_1, M_2, M_3) is used to make reconstructions over a kappa range $\kappa = [1.5; 3.5]$ in steps of 0.2 based on the five CM skull representations with different amounts of inserted outliers.

Figure 7.14 depicts the overall global rmse over 393 cases as a function of kappa similar to figure 7.4 and 7.10. Based on these plots we select a single kappa value for every craniofacial model having the lowest overall global rmse, resulting in a representative reconstruction for every model and noisy skull representation with a ceratin amount of outliers. Doing so, we can construct figures 7.15 and 7.16 similar to figures 7.11 and 7.12 to compare the different craniofacial models and to validate the contribution of the EM skull registration framework in the case of outliers present in the skull representation. A first observation is the robustness to outliers for the four different models embedded into the EM framework. Especially when observing the recognition performance curves in figure 7.15. Secondly, according to figure 7.14, a high kappa value is preferred when the amount of outliers is limited, because the majority of the skull data gives a proper noise level estimation of 1mm according to which the few outliers can be easily detected and rejected even with a high kappa value. A kappa value which is too low discards non-outlier or inlier

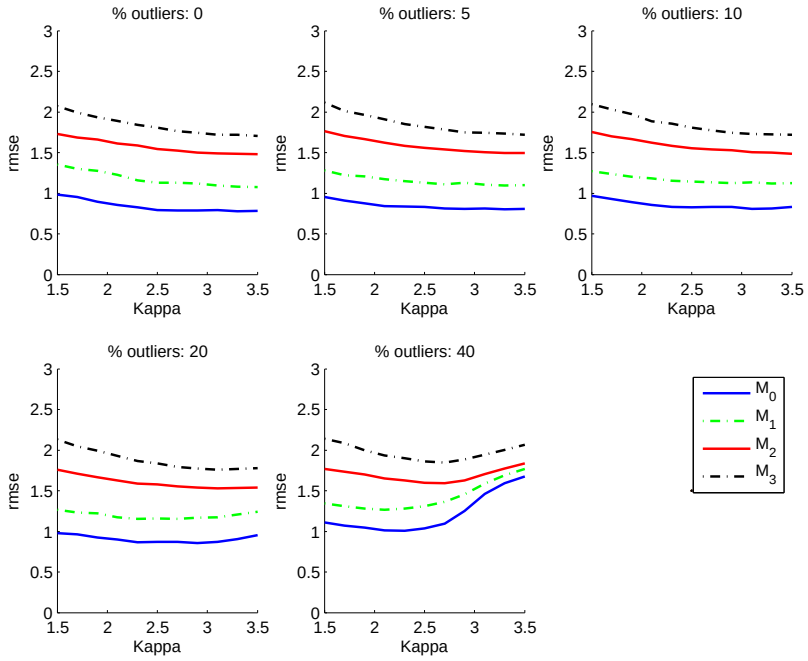


Figure 7.14: The overall global rmse results for five different amounts of outliers in a CM skull representation in function of kappa. Vertical axes: overall global rmse over 393 cases for every model. Horizontal axes: kappa value.

skull landmarks. Increasing the amount of outliers results into a lower preferred kappa value, because fewer skull landmarks are trustworthy to estimate the proper amount of noise according to which the outliers are to be detected.

Concentrating onto the statistical models M_0 and M_1 in figure 7.14, we observe that the increase of rmse for M_0 is more severe with increasing kappa value and an increase of the amount of outliers. A skull landmark displacement has mainly an influence on the first condition in the craniometric skull similarity criterium, resulting in a more difficult correct thickness estimation. This thickness relationship based on an outlier, influences severely both thickness and surface estimates within M_0, while the influence is mostly limited to the thickness estimate within M_1. The same difference in kappa sensitivity was observed using the implicit skull representation in the clinical patient database, which was characterized by an amount of outliers present in the VIS/ultrasound thickness value relationship. Finally, observing figure 7.16, we see the same improvements as in figure 7.12 using the proposed EM skull registration framework for every craniofacial model, but now in terms of outlier robustness. The positive contribution of the robust statistical estimation (E-step during EM) is more pronounced compared to the noisy skull representations in figure 7.12, because severe outliers were now explicitly incorporated.

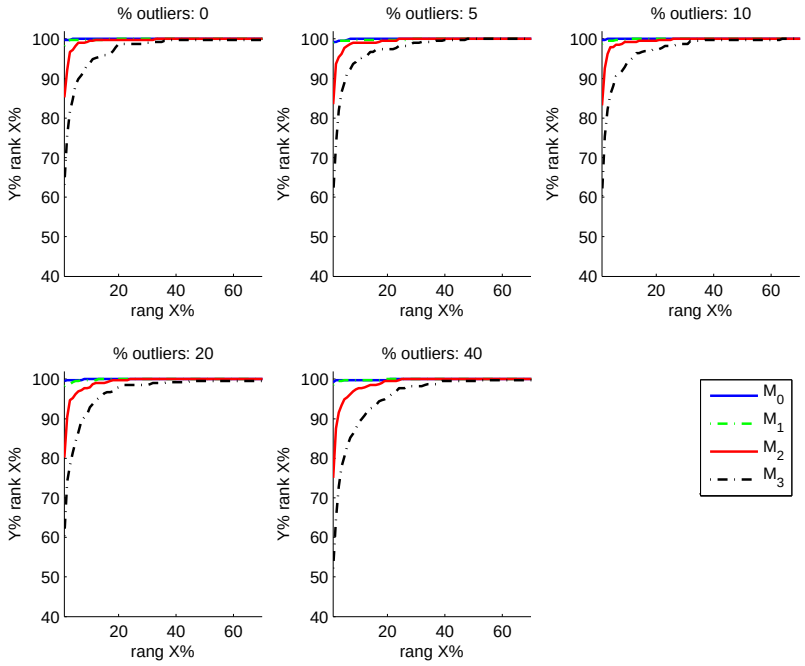


Figure 7.15: Cumulative recognition curves for 393 cases using the four different craniofacial models based on five different amounts of outliers in the skull representation

7.5.4 Forensic and Archeological cases

The primary application for our craniofacial reconstruction methodology is victim identification. Besides forensic related reconstructions, archeological facial reconstructions could also be tackled with our methodology. In this section both a forensic and a archeological attempted craniofacial reconstruction example are given as final teasers before we conclude this chapter in the next section.

Forensic case

The forensic related reconstruction concerns a skull found in Belgium in April 2006. The anthropological examination estimated the skull gender to be female and the age to be 45 years old. Furthermore a BMI of 22.5 (slender woman) was estimated. Based on the validation results from the previous two sections, our primary craniofacial model choice is the model defined in chapter 6, while using (7.15), (7.16), (7.17) and (7.18). Our primary skull representation choice is the implicit representation defined in section 7.4.1, generating a complete automatic craniofacial reconstruction while being acceptable in terms of accuracy and recognition. The user-specific amount of regularization was set

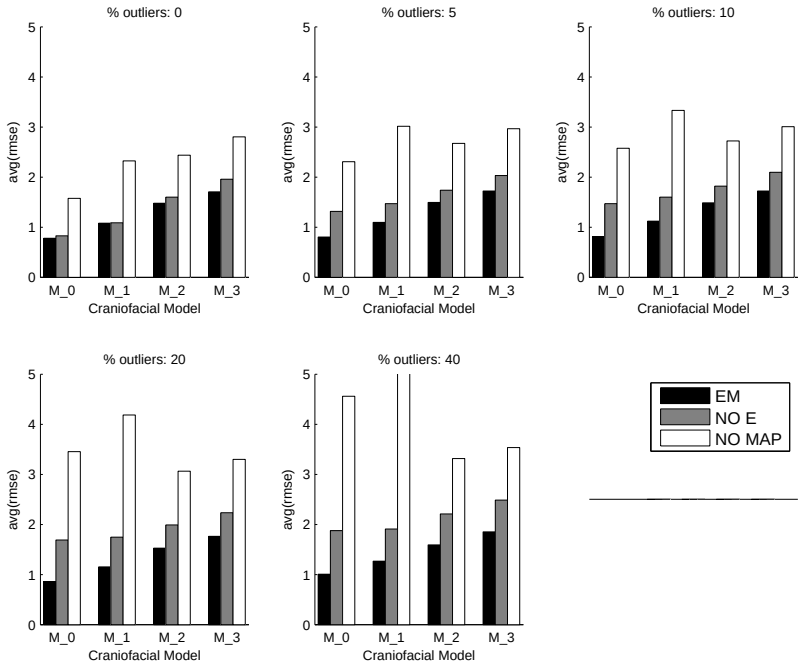


Figure 7.16: Overall global rmse over 393 cases of the representative reconstructions using the EM skull registration framework and the best kappa value, Without the E-step or outlier robustness and without the MAP formulation or noise robustness for every craniofacial model (M_0, M_1, M_2, M_3) and five CM skull representations with different amounts of outliers

equal to one. $\nu = 1$. Based on the kappa analysis in figure 7.4 we selected a kappa value $\kappa = 2$. The reconstruction result based on the anthropological examination is given in figure 7.17. Alternative property values are depicted in figure 7.18.

Archeological case

The archeological reconstruction concerns the skull of a Neanderthaler [105], illustrated in figure 7.19. A virtual reassembling of the skull based on skull fragments was accomplished by Prof. Dr. C.P.E. Zollikofer and Dr. M. Ponce de Len from the University of Zürich. Furthermore, they indicated the 52 skull landmarks on the skull according to the protocol given in appendix F. An exchange of the complete skull data was infeasible due to contractual constraints such that the skull representation choice is a craniometric representation based on the indicated skull landmarks. We used the statistical craniofacial model from chapter 6 and selected a kappa value $\kappa = 3$. In a first instance the user-specific amount of regularization was set equal to one, resulting into the

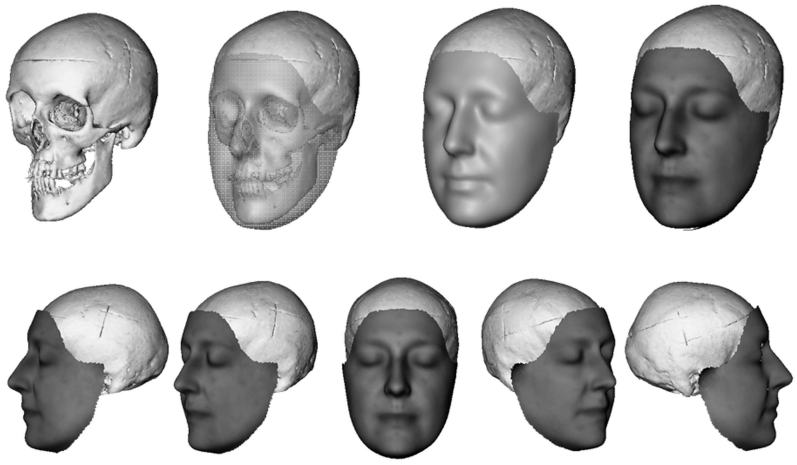


Figure 7.17: Forensic related real reconstruction case of a skull found in Belgium

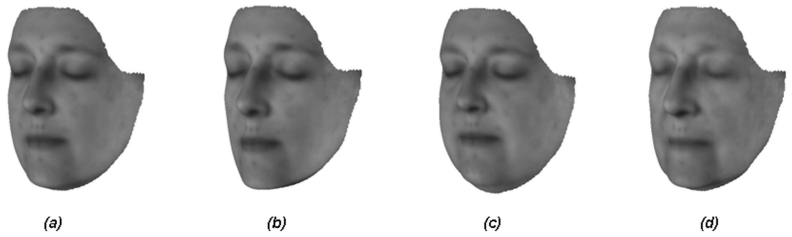


Figure 7.18: Alternative property choices for the forensic case. (a) original from figure 7.17 (b) BMI = 17 (c) BMI = 28 (d) Age = 80

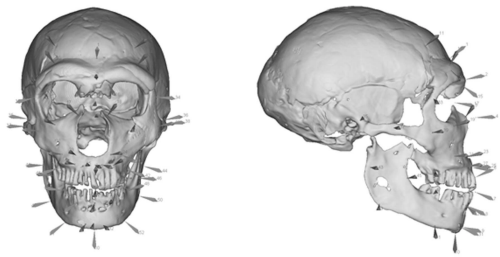


Figure 7.19: Illustration of the Neanderthaler skull with indicated skull landmarks

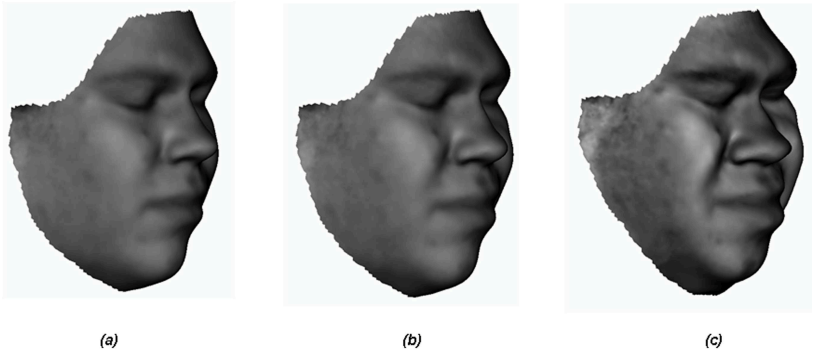


Figure 7.20: Neanderthaler reconstructions based on the skull landmarks with varying user-specific regularization. (a) $\nu = 1$ (b) $\nu = 0.1$ (c) $\nu = 0.05$

reconstruction of figure 7.20(a). It is observed that the final outlook resembles the face of a modern man, because the database, used to construct the craniofacial model, contains facial information from living subjects of today instead of Neanderthaler specimens. Setting $\nu = 1$ constrains the reconstruction to stay within the statistical boundaries of a modern population. In other words, a trade-off between modern facial plausibility and skull data fidelity is present, favoring plausible solutions. By lowering the user-specific amount of regularization we obtain the reconstruction results in 7.20(a) and (b) for $\nu = 0.1$ and $\nu = 0.01$ respectively. Doing so, the trade-off is changed in favor of the skull data fidelity and it is observed that the reconstructions tend towards the facial outlook of a (possibly) Neanderthaler. Lowering the statistical boundary constraints or the statistical regularization allows the reconstruction of faces a-typical to the database. Very speculatively, we can state that the three reconstructions in figure 7.20 show an evolution of a modern man towards a Neanderthaler and vice-versa. While the reconstruction in 7.20(c) shows the Neanderthaler outlook, the reconstruction in 7.20(a) shows the outlook of the modern man version based on the Neanderthaler skull. The facial outlook in figure 7.20(b) is a version in between the stone-age Era and the modern-age Era.

7.6 Conclusion

This chapter presented the different components needed for craniofacial reconstruction. We proposed a statistical reconstruction approach: *What is the most plausible or probable face, according to a craniofacial model, given the (erroneous) skull data ?*. In order to find the most plausible face a probabilistic EM skull registration framework applicable to a wide range of craniofacial model and skull representation choices was given. The choice of craniofacial model mainly influences the transformation model. Furthermore, the model

template influences the amount of floating surface points used during registration and the predefined distances, depending on the craniofacial relating information contained within the database from which the model was deducted. We defined and proposed two statistical based and two TPS based craniofacial model choices deducted from our database, starting with the model from chapter 6 and by subsequently eliminating covariance knowledge. The choice of skull representation determined the current distance computations and the individual point similarity measures, therefore influencing the inlier-process component within the data-likelihood term. The first and primary skull representation choice that we proposed was an implicit skull representation. The second choice was a more traditional craniometric skull representation based on manually indicated craniometric or anatomical skull landmarks.

A first set of validation tests were performed based on a clinical patient database, simulating realistic reconstruction scenarios. Two craniometric skull representations of different quality were used in addition with an implicit skull representation based on variational implicit surfaces per case in the database. A second set of validation tests were performed based on the model database, through a leave-one-out procedure, simulating controlled noisy and outlier skull data points in a craniometric representation. Reconstruction evaluations were obtained based on a quantitative and a qualitative comparison of the reconstructed facial surfaces with the ground truth facial surfaces. The quantitative measures were a global and a local rmse of the remaining distances between corresponding points on the surfaces of the reconstruction and the ground truth. The qualitative measure was a recognition or identification success rank obtained by comparing the reconstruction with faces in a database including the ground truth using a similarity measure or distance between faces using the cosine of the angle between re-scaled model descriptions.

Based on the validation tests a few conclusions could be drawn concerning the craniofacial models incorporating different kinds of covariance information. Firstly, a clear improvement in reconstruction performance was obtained simulating and eliminating property related variance, restricting the facial surface and thickness variation according to a given set of skull properties derived from an anthropological examination. Secondly, a major improvement in reconstruction performance was gained using face-specific deformations extracted from facial surface and landmark covariance. Statistical knowledge or internal structural relationships of multiple faces was used instead of a single internal structural relationship to deform the averaged model template. Thirdly, a varying improvement could be obtained incorporating thickness covariance, depending on the skull representation. Based on a noiseless and outlier free skull representation, the co-varying knowledge between thicknesses and facial surfaces improved the reconstructions significantly. The improvement was less significant but still present when the amount of severe outliers was increased. However, a proper choice of kappa was necessary, because outliers had an increased influence on both thickness and facial surface estimations compared to the independent co-varying thicknesses and facial surfaces. Increasing the noise level in the skull representations lead to comparable results between the

models with and without thickness covariance. However, the choice of kappa became more important for the statistical model without thickness covariance, because extra fine-tuning was needed to restrict the independent flexibility between thicknesses and surfaces in a noisy setup. Incorporation of thickness covariance was interesting when the link between the skull representation and the model expressed in the similarity criterium was strong. Increasing the noise level weakened the link, such that the contribution of thickness covariance was reduced, until the noise level was too high and no more positive contribution was observed. Incorporating thickness covariance in the presence of outliers was useful if the outliers were properly detected and down-weighted.

The semi-automatic results based on the craniometric skull representation were dependent on the quality of the manually indicated landmarks, related with the expertise of the anthropologist. A realistic indication without knowledge of the ground truth facial surface was very noisy and error prone, favoring slightly the craniofacial model without thickness covariance. A good skull landmark indication lead to flawless recognition results independent of kappa using the more statistical based craniofacial models. A proper anthropological training was therefore advised in order to improve the skull landmark indications and as such the reconstructions. Complete automatic reconstruction results, eliminating human subjectivity, were obtained using the implicit skull representation. This representation was characterized less noisy than the realistic craniometric representation but a severe amount of outlier VIS/ultrasound value relationships were present. Therefore, better reconstruction results were obtained using a lower kappa value. The proposed statistical craniofacial model was slightly favored with this skull representation compared to the statistical model without thickness covariance, but the dependency on kappa was more pronounced because of the increased influence of outliers through the use of thickness covariance. Finally, the validation tests clearly showed an improved robustness against noise and outliers within the skull representation using the proposed robust EM skull registration framework. The MAP formulation lead to robustness against noise or small errors, while the further embedding into an EM formulation of the registration problem leaded to robustness against outliers or gross errors as well.

In the following chapter we give a general conclusion and some future work issues concerning the methodologies proposed and developed in this thesis.

Chapter 8

Conclusions and future directions

8.1 Craniofacial reconstruction

When confronted with a corpse that is unrecognizable due to its state of decomposition, skeletonisation, mutilation or incineration, and if no other identification evidence is available, craniofacial reconstruction could be considered to help the investigation out of an impasse. Forensic facial reconstruction aims at estimating the facial outlook associated to an unknown skull specimen for victim identification. All facial reconstruction techniques were based on the assumed relationship between the soft tissue envelope and the underlying skull substrate. Manual reconstruction methods required a lot of anatomical and artistic modeling expertise and are as a result subjective and time consuming. The development of software for computerized facial reconstructions of an individual would be of benefit to various law enforcement agencies, by allowing faster, easier and more efficient generation of multiple representations of an individual.

8.1.1 Computerized reconstruction

Computer-based methods were essentially a virtual mimicking of manual reconstruction techniques. A general work-flow for computerized techniques could be defined involving an anthropological examination influencing the choice of craniofacial model. Subsequently this model was deformed or transformed towards a virtual copy of the skull based on a similarity criterium extracted from a skull representation. The proper transformation parameters were determined by optimizing an objective function combining the similarity criterium and the craniofacial transformation model.

Besides the issues about craniofacial reconstruction in general, a number of choices were to be made and a number of implementation issues were to

be dealt with while designing a computerized reconstruction technique. For every step in the general work-flow several choices exist and problems were to be coped with.

8.1.2 Our approach

We proposed a complete statistical approach for computer-aided craniofacial modeling and reconstruction. The philosophy behind the reconstruction technique was to find the most plausible or probable face, according to a face distribution or model, given the erroneous skull data. A statistical craniofacial model for reconstruction was used, eliminating the model-bias and minimizing the unrealistic character of the reconstructions caused by large generic model deformations, incorporating the combined population-dependent variance and covariance of complete skin surface shape, 52 anatomical face landmarks with ultrasound tissue depths, property (age, BMI and gender) values and skin surface gray-value texture information, calculated from an elaborate database of faces measured in an upright position. Furthermore, statistical manipulation or simulation of properties was used to fine-tune the model towards a given set of skull properties. The model could be considered as an elastic mask with elastic hemi-spherical dowels on the inside of the mask at the landmark locations, which was subsequently fitted to the skull such that the virtual dowels touched the skull. The elastic deformation of the transformation model was based on the statistical likelihood as learned from a database of nearly 400 subjects.

In order to build up the statistical craniofacial model and to use it for craniofacial reconstruction several registration subtasks were to be tackled. (1) The acquisition of 3D skin surface and tissue-depth information in an upright position, measured over a sufficiently large and diverse population which were subsequently stored in a database together with the subject's age, Body Mass Index (BMI), gender and ancestry information. (2) The establishment of inter-subject surface dens correspondences to perform statistical modeling. (3) Statistical modeling of the population-dependent variation and covariation of skin surface shape and tissue depth, fine-tuned towards a given set of skull properties (ancestry, age,...). (4) Fitting of this statistical model to the individual craniofacial skeleton. (5) Validation of the complete craniofacial reconstruction approach.

The whole procedure (1-5) involved a number of registration undertakings. First in subtask (1), complete facial surfaces were assembled from partial surface patches using a rigid transformation model. Second in subtask (2), surfaces of different subjects were matched to each other using a non-rigid Thin Plate Spline based deformation model. In subtask (3), the observed inter-individual deformations were statistically modeled via Principal Component Analysis, assuming a multivariate normal distribution. This model was matched in a first instance to the facial surfaces of the different subjects in the database using a transformation model based on the learned statistical deformations to analyze the quality of the Thin Plate Spline based registrations

in the previous subtask. Finally in subtask (4), this facial model (including the soft-tissue thicknesses) was aligned with the (virtual) skull specimen using again the transformation model based on the learned statistical deformations.

8.2 Methods

8.2.1 Robust surface registration framework

Surface registration referred to the establishment of the geometrical relationship between two or more surfaces. We presented a robust, statistical, surface registration framework that optimized an inter-surface distance measure over the parameters that defined the geometric transformation. We used a memory efficient, continuous, smooth and analytical, implicit surface representation, which has particularly interesting properties for registration purposes. In addition, a variational boundary function to retain the boundary concept within an implicit surface representation for open surfaces was created.

Components generally needed for registration were introduced and elaborated while focussing on the use of implicit surfaces and a robust statistical objective function. We proposed to use a statistical Maximum a Posteriori (MAP) derivation of an objective function to combine a geometric-based similarity measure and the transformation regularization. MAP approaches have proven to be very useful theoretical frameworks, equipped with a set of powerful and well developed mathematical techniques having a sound philosophical underpinning. Furthermore, all the assumptions were clear and on the table. Robustness against noise was easily obtained by stating that the similarity measure instances (for a certain set of transformation parameters) within the floating surface points were generated by a statistical modeled noisy inlier-process. Robustness against outliers was obtained by stating that the same similarity measure instances or samples were either or partially generated by a probabilistic inlier-process or/and an outlier-process constituting a complete-process. Doing so, an outlier latent variable was introduced encoding the amount of inlier-belief and outlier-belief for a similarity measure value in a floating surface point and was therefore a kind of confidentiality variable. Four modeling choices for the outlier latent variable were given dependent on the prior knowledge concerning the outlier generating process. The complete-process modeling leaded to a joint optimization, similar to the EM algorithm, of transformation parameters and outlier latent variables.

Equivalent M-estimators after outlier latent variable elimination were constructed for the different kinds of latent variable choices. Doing, so we were able to analyze outlier influences. The construction of the M-estimators was obtained during the E-step of the joint estimation regarded as a general interpretable EM optimization. Thanks to the implicit surface representations no additional correspondence search was needed and a straightforward Steepest-Descent M-step optimization could be applied. Furthermore, well established optimization routines for robust statistics were directly applicable on the ex-

plicit outlier-process formulations due to the unification with robust M-estimators. The final result was a better underpinned objective function with quasi self-regulating robustness against noise and outliers.

Future directions concerning the robust statistical surface registration framework include the incorporation of additional surface information besides geometry. Using the same VIF machinery, texture and curvature information on the target surface can be extrapolated towards the floating surface, such that the similarity measure can contain texture and curvature match evaluations as well. The surface registration framework was applied to tackle the patch, complete facial surface and skull registration tasks, but we are sure that other registration undertakings can be solved using the framework.

8.2.2 Facial surface acquisition

Facial surface entries within the craniofacial database were acquired using a photogrammetric active 3D acquisition device. Advantages included, portability and upright position acquisition. However, due to the limited viewing-angle of the camera a complete facial surface was assembled from multiple partial surface acquisitions. To build a complete database it was interesting to develop an automatic assembling procedure for the facial surface acquisition.

We presented an automatic, robust and accurate registration and integration algorithm for 3D surface building from separate patches, without any prior knowledge. We used a variational implicit surface representation of the patches. The VIS representation had useful properties for solving the different tasks and had interesting smoothing possessions to deal with noisy data. First, a pair-wise crude registration was solved by making use of directed point-signatures in combination with a voting scheme. Shortcomings of spin-images for use as point-signatures were solved by creating VISS-images increasing the quality and the robustness of the point-signatures. A MST was used to convert the pair-wise crude registrations into a multi-view crude registration resulting in a set of best transformations, according to the number of transformation inliers, needed to transform all the patches into a common coordinate system. This set of best transformations was further refined by applying pair-wise fine registrations. We defined an objective function according to the surface registration framework, based on the VIS representation of the target patch. The optimization was accomplished by a six dimensional conjugate gradient search framework according to the transformation parameters, instead of finding corresponding points between floating patch and target patch. We also defined a novel variational implicit function for an improved outlier detection and rejection during optimization. In contrast to traditional outlier detection systems, the outlier detection function depended on the target patch only, resulting in an easier outlier threshold setting. Finally, accumulation of pair-wise registration errors was dealt with by applying a multi-view simultaneous fine registration. Solving this registration was similar to the pair-wise fine registration except that the target patch was replaced by an intermediate volumetric integration of patches. The integration itself was also based on the VIS representations

such that integration and registration could be unified.

By making use of variational implicit surface and function representations of the patch surfaces, textures and weights we solved limitations of current volumetric integration methods. No memory inefficient 3D grid was used, such that the final resolution or accuracy was not limited by memory requirements. Hole filling did not require an extra step because of the interpolation and extrapolation properties of the chosen representations. Furthermore, the representation had useful smoothing properties when dealing with noisy input data. Texture integration was performed similarly to shape integration, making use of the same weight functions in order to reduce the texture influence of low confidence patch points.

We performed experiments on real-live data, noiseless and noisy simulated data. The real-live data set showed the robustness of the algorithm to bad data. In order to improve the final quality of the 3D model, some really bad data parts in the patches were to be deleted or more patches are to be taken in order to reduce the effect of the bad parts in the final integration. The noiseless data set showed the accuracy of the algorithm. The noisy simulated data set illustrated that the smoothing properties of the VIS representation can be used to make the algorithm robust against noisy input data.

The proposed algorithm is based on geometric properties of the patches only. Future research will also focus on incorporating other information, like texture, into the different steps of the algorithm. Furthermore, other and better confidence definitions for patch points to cope with bad data during integration are to be explored.

8.2.3 Facial surface registration

The last prerequisite for our craniofacial database was to obtain registered inter-subject facial surfaces to construct a statistical craniofacial model. Therefore, the 3D points with known connectivity of a carefully constructed generic reference face were mapped or registered non-rigidly onto the faces in the database. The facial surface registration methodology had to be robust against noise and had to be be able to cope with outliers due to partial overlap or missing data. Therefore we reformulated the registration problem within the prosed robust surface registration framework. The reference face was considered to be the floating surface represented as a set of floating points, while a face from the database was considered to be the target surface represented with a VIS such that no sampling errors occurred. A proper non-rigid TPS based transformation model and similarity measure, based on the VIS target surface representation, were chosen and both were combined into a single objective function. To handle a vast amount of floating surface points, we used the VIF theory and algorithmic improvements to update and evaluate the transformation. Outliers were dealt with using an outlier detection function equal to the one used during patch registration, incorporating strong outlier prior knowledge. A deterministic annealing optimization strategy while alternating between inlier, point correspondence and transformation updates was applied

to cope with the non-convex nature of the objective function.

Some results were shown to perform a validation concerning accuracy and consistency. The former was easy to analyze and was acceptable. The latter was more difficult to analyze, but already some visual consistency results were shown and were promising. Further consistency validation, with satisfying results, was obtained using the model derived from the TPS based correspondences. The statistical model, being an elastic mask, was embedded within the proposed surface registration framework to tackle a model-based facial surface registration problem and the results were compared to the TPS based approach for unseen facial surfaces.

Similar to the patch registration, future research will focus on the incorporating of other facial information, like texture and curvature, using the VIF machinery. Furthermore, a simultaneous TPS based correspondence establishment and model building is thinkable to obtain a better final model quality. After a first run of TPS based registrations an initial model, incorporating limited overall database variance information, can be constructed and the model based facial surface registration approach can be used to obtain a facial entry approximation. This approximation is then used to initialize (better than then the average face) a second TPS based registration run for every facial entry in the database. The whole process can be repeated with increasing amount of overall variance captured in the model to further initialize an additional TPS based registration run, until the initializations are accurate enough without an additional TPS based registration.

8.2.4 Statistical craniofacial model

We proposed a statistical craniofacial model for skull reconstruction purposes, modeling the combined population dependent variance and covariance of skin surface shape, texture, tissue depths and properties. The model was constructed, based on a database consisting of 393 individuals, using principal component analysis generating an average face, being the model template, combined with a set of face-specific deformations or principal components, being the transformation model. The model capacity was examined for in-training and unseen faces and appeared to be sufficient in terms of incorporated percentage variance and number of training samples.

The statistical model, being an elastic mask, was embedded within the surface registration framework to tackle a model-based facial surface registration problem and was compared to the TPS based approach. Furthermore a slight preview concerning skull instead of facial surface registration was given by using the sparse face landmarks instead of dens surface points as floating surface points. This showed an appropriate strength of the co-varying facial information knowledge incorporated in the eigenvector matrix to solve skull registrations based on a limited amount of registration evidence.

An additional prerequisite for a model being a valid craniofacial model was the ability to manipulate or alternate the model based on a given skull property set obtained from an anthropological examination. Instead of selecting a sub-

population from the database we opted to learn and to remove facial variations from the model originating from differences in property values. The result was a flexible (containing enough variation) but property normalized statistical craniofacial model such that skull properties are obeyed as a hard constraint prior to skull registration. The incorporation of gray-valued texture variation generated a more lifelike reconstruction based on a texture map deducted from facial geometry, property value and tissue depth information. We believed that using gray-valued instead of colored texture information was less dangerous to trigger the recognition process into the wrong direction.

Future directions includes the use of alternative statistical modeling techniques besides principal component analysis, not necessarily assuming a multivariate normal distribution of the facial data. Explorations consists of Partial Least Square (PLS) models and non-linear based statistical models. Although for the latter a better distributed database is required. Furthermore, ageing is known to be a non-linear process. Therefore, it would be interesting to expand the database with facial entries over a diverse age range and to model the ageing non-linearly, while comparing it with the linear age simulation.

8.2.5 Skull registration

We proposed a statistical reconstruction approach: *What is the most plausible or probable face, according to a craniofacial model, given the (erroneous) skull data ?*. In order to find the most plausible face a probabilistic EM skull registration framework, based on the presented surface registration framework, applicable to a wide range of craniofacial model and skull representation choices was given. The choice of craniofacial model mainly influenced the transformation model. Furthermore, the model template influenced the amount of floating surface points used during registration and the predefined distances, depending on the craniofacial relating information contained within the database from which the model was deducted. We defined and elaborated two statistical based and two TPS based craniofacial model choices deducted from our database, starting with the proposed statistical model and by subsequently eliminating covariance knowledge. The choice of skull representation determined the current distance computations and the individual point similarity measures, therefore influencing the inlier-process component within the data-likelihood term. The first and primary skull representation choice that we proposed was an implicit skull representation. The second choice was a more traditional craniometric skull representation based on manually indicated craniometric or anatomical skull landmarks.

A first set of validation tests were performed based on a clinical patient database, simulating realistic reconstruction scenarios. Two craniometric skull representations of different quality were used in addition with an implicit skull representation based on variational implicit surfaces per case in the database. A second set of validation tests were performed based on the model database, through a leave-one-out procedure, simulating controlled noisy and outlier skull data points in a craniometric representation. Reconstruction evaluations

were obtained based on a quantitative and a qualitative comparison of the reconstructed facial surfaces with the ground truth facial surfaces. The quantitative measures were a global and a local rmse of the remaining distances between corresponding points on the surfaces of the reconstruction and the ground truth. The qualitative measure was a recognition or identification success rank obtained by comparing the reconstruction with faces in a database including the ground truth using a similarity measure or distance between faces using the cosine of the angle between re-scaled model descriptions.

Based on the validation tests a few conclusions could be drawn concerning the craniofacial models incorporating different kinds of covariance information. Firstly, a clear improvement in reconstruction performance was obtained simulating and eliminating property related variance, restricting the facial surface and thickness variation according to a given set of skull properties deducted from an anthropological examination. Secondly, a major improvement in reconstruction performance was gained using face-specific deformations extracted from facial surface and landmark covariance. Statistical knowledge or internal structural relationships of multiple faces was used instead of a single internal structural relationship to deform the averaged model template. Thirdly, a varying improvement could be obtained incorporating thickness covariance, depending on the skull representation. Based on a noiseless and outlier free skull representation, the co-varying knowledge between thicknesses and facial surfaces improved the reconstructions significantly. The improvement was less significant but still present when the amount of severe outliers was increased. However, a proper choice of kappa was necessary, because outliers had an increased influence on both thickness and facial surface estimations compared to the independent co-varying thicknesses and facial surfaces. Increasing the noise level in the skull representations leaded to comparable results between the models with and without thickness covariance. However, the choice of kappa became more important for the statistical model without thickness covariance, because extra fine-tuning was needed to restrict the independent flexibility between thicknesses and surfaces in a noisy setup. Incorporation of thickness covariance was interesting when the link between the skull representation and the model expressed in the similarity criterium was strong. Increasing the noise level weakened the link, such that the contribution of thickness covariance was reduced, until the noise level was to high and no more positive contribution was observed. Incorporating thickness covariance in the presence of outliers was useful if the outliers were properly detected and down-weighted.

The semi-automatic results based on the craniometric skull representation were dependent on the quality of the manually indicated landmarks, related with the expertise of the anthropologist. A realistic indication without knowledge of the ground truth facial surface was very noisy and error prone, favoring slightly the craniofacial model without thickness covariance. A good skull landmark indication leaded to flawless recognition results independent of kappa using the more statistical based craniofacial models. A proper anthropological training was therefore advised in order to improve the skull landmark

indications and as such the reconstructions. Complete automatic reconstruction results, eliminating human subjectivity, were obtained using the implicit skull representation. This representation was characterized less noisy than the realistic craniometric representation but a severe amount of outlier VIS/ultrasound value relationships were present. Therefore, better reconstruction results were obtained using a lower kappa value. The proposed statistical craniofacial model was slightly favored with this skull representation compared to the statistical model without thickness covariance, but the dependency on kappa was more pronounced because of the increased influence of outliers through the use of thickness covariance. Finally, the validation tests clearly showed an improved robustness against noise and outliers within the skull representation using the proposed robust EM skull registration framework. The MAP formulation leaded to robustness against noise or small errors, while the further embedding into an EM formulation of the registration problem leaded to robustness against outliers or gross errors as well.

Future directions mainly concentrate on the improvement or replacement of the VIS/ultrasound relationship for the implicit skull representation. Instead of using ultrasound thicknesses we could use a CT based facial database with complete skull surface information. The ultrasound thicknesses can then be replaced by VIS values in the face landmarks calculated from the VIS of the complete skull surfaces in the database. The craniofacial model can then use a VIS/VIS value relationship instead, which will result in lesser outliers during registration and thus more information to guide the reconstruction. Furthermore we will not be limited to 52 face landmarks, because VIS values can be automatically calculated for a larger subset of facial surface points. Additionally, care must be taken to set-up a proper face-pool validation framework. A more realistic, human subjective, identification process can be simulated by generating face-pool tests. Given an image of a reconstruction and a set of possible candidate images extracted from the database, a human observer is asked to indicate the face from the face-pool most similar to the given reconstruction. However, it is very difficult to generate realistic face-pool tests. Finally, a proper influence analysis of erroneous skull property predictions is to be performed.

8.3 The end

Craniofacial reconstruction balances between the world of Art and Science. Through the efforts of many people around the world the line is slowly but surely shifted towards science to improve the practical relevance of craniofacial reconstructions within crime-scene investigations and justice convictions. I'm convinced that the value of craniofacial reconstructions can largely be improved, when combined with appropriate facial recognition and identification techniques, which is an active research topic as we speak.

Bibliography

[1] S. De Greef, P. Claes, W. Mollemans, M. Loubele, D. Vandermeulen, P. Suetens, and G. Willems. Semi-automated ultrasound facial soft tissue depth registration : method and validation. *Journal of Forensic Science*, 2005. 7, 29, 125, 126, 229

[2] S. De Greef, P. Claes, D. Vandermeulen, W. Mollemans, P. Suetens, and G. Willems. Large-scale in-vivo caucasian facial soft tissue thickness database for craniofacial reconstruction. *Forensic science international*, 159(1):126–146, 2006. in press. 7, 29, 125, 126

[3] P. Claes, S. Degreef, G. Willems, D. Vandermeulen, and P. Suetens. Craniofacial statistical modeling and reconstruction. In *3D modelling*, April 28-29 2004. Paris,France. 11, 27, 29, 32, 34, 35

[4] P. Claes, D. Vandermeulen, S. De Greef, G. Willems, and P. Suetens. Combined statistical modeling of tissue depth and 3d facial outlook for computerized facial approximation. In *2nd international conference on reconstruction of soft facial parts*. RSFP, March 17-18 2005. Remagen, Germany. 11, 27, 28, 29, 33, 34, 35

[5] P. Claes, D. Vandermeulen, S. De Greef, G. Willems, and P. Suetens. Statistically deformable face models for cranio-facial reconstruction. In *Proceedings 4th international symposium on image and signal processing and analysis*, pages 347–352. ISPA, September 15-17 2005. Zagreb,Croatia. 11, 27, 29, 33, 34, 35

[6] P. Claes, D. Vandermeulen, S. De Greef, G. Willems, and P. Suetens. Craniofacial reconstruction using a combined statistical model of face shape and soft tissue-depths : methodology and validation. *Forensic science international*, 159(1):147–158, 2006. 11, 27, 28, 29, 33, 34, 35, 39

[7] P. Claes, D. Vandermeulen, S. De Greef, G. Willems, and P. Suetens. Statistically deformable face models for cranio-facial reconstruction. *Journal of computing and information technology*, 14(1):21–30, 2006. 11, 27, 28, 29, 33, 34, 35, 39, 248, 249

[8] K.T. Taylor. *Forensic art and illustration*. New York: CRC Press, 2001. 21

[9] M. Gerasimov. The face finder, 1971. Philadelphia, PA: JB. Lippencott Co. 21

[10] G.U. Lebedinskaya, T.S. Balueva, and E.B. Veselovskaya. *Development of methodological principles for reconstruction of teh face on the basis of skull material. In Iscan M.Y. and Helmer R.P., Forensic analysis of teh skull.* Wiley-Liss Inc.,New York, 1993. 21

[11] C.C. Snow, B.P. Gatliff, and K.R. Mc Williams. Reconstruction of the facial features from skull: An evaluation of its usefulness in forensic anthropology. *American Journal of Physical Anthropology*, 33:221–228, 1970. 21

[12] J. Prag and R.A.H Neave. *Making Faces using forensic and Archeological evidence.* British Museum Press, 1997. London. 22

[13] C. Wilkinson. *Forensic Facial Reconstruction.* Cambridge University Press, Cambridge, 2001. 22

[14] S. Davy, T. Gilbert, D. Schofield, and M.P. Evison. *Forensic facial reconstruction using computer modeling software, in: J.G Clement, M.K. Marks, Computer-Graphic Facial Reconstruction*, chapter 10, pages 183–196. Elsevier Academic Press, 2005. 22, 24, 26, 32, 34, 38

[15] W.D. Haglund and D.T. Reay. Use of facial approximation techniques in identification of green river serial murder victims. *American Journal of Forensic Medicine and Pathology*, 12(2):132–142, 1991. 22

[16] J.B. Schimmler, R.P. Helmer, and J. Rieger. *Craniometric individuality of human skulls, in M.Y, Iscan and R.P. Helmer, Forensic analysis of the skull*, pages 89–96. New York: Wiley-Liss, Inc, 1993. 24

[17] K.J. Reichs. Forensic anthropology in the 1990s. *The American Journal of forensic medicine and pathology*, 13(2):146–153, 1992. 24

[18] S.D. Ousley and R. Jantz. Fordisc 2.0: Personal computer forensic discriminant functions. Technical report, University of Tennessee, Knoxville, Tennessee, 1996. 24

[19] D.H. Ubelacker. Fordisc 2.0: Personal computer forensic discriminant functions. *International Journal of Osteoachaology*, 8:128–133, 1998. 24

[20] S.D. Ousley and R. Jantz. Fordisc 3.0: Personal computer forensic discriminant functions. Technical report, University of Tennessee, Knoxville, Tennessee, 2005. 24

[21] A.I.M. Shaweesh, J.G. Clement, C.D.L. Thomas, and A. Bankier. Construction and use of facial archetypes in anthropology and syndrome diagnosis. *Forensic Science International*, 159(1):175–185, 2006. 24

[22] P. Vanezis. Application of 3-d computer graphics for facial reconstruction and comparison with sculpting techniques. *Forensic Science International*, 42:69–84, 1989. 24, 26, 29, 32, 34

[23] A.W. Shahrom, P. Vanezis, R.C. Chapman, A. Gonzales, C. Blenkinsop, and M.L. Rossi. Techniques in facial identification : computer-aided facial reconstruction using laser scanner and video superimposition. *Journal of Legal Medecine*, 108:194–200, 1996. 24, 26, 29, 34, 35

[24] M.P. Evison. Computerised 3d facial reconstruction. In *REsearch school of archeology and archeological science*, 1996. University of Sheffield, UK. 24, 27, 28

[25] M.P. Evison. Modeling age, obesity, and ethnicity in a computerized 3-d facial reconstruction. In *Proc. 9th Biennial meeting of the international association for craniofacial identification*. FBI, July 2000. Washington,DC. 24, 27, 28

[26] B. Lorensen and H.E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proc. of ACM Siggraph*, volume 21, pages 163–169, July 1987. 24, 47

[27] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, 2002. 24

[28] F. Seibert. Model-based 3d reconstruction of human faces. In *Computer Graphic Topics*, volume 3, pages 8–9, 1997. 24, 27, 28, 29, 32, 34, 36

[29] R. Evenhouse, M. Rasmussen, and L. Sadler. Computer-aided forensic facial reconstruction. *Journal of BioCommunication*, 19(2):22–28, 1992. 26, 29, 32, 34, 35

[30] B. Andersson and M. Valfridsson. Digital 3d facial reconstruction based on computer tomography. Master's thesis, Linkopings university, Sweden, February 2005. 26, 29, 34

[31] K.M. Archer. Cranio-facial reconstruction using hierarchical b-spline interpolation. Master's thesis, The University of Britisch Colombia, 1997. 26, 29, 32, 34, 35

[32] K. Archer, K. Coughlan, D. Forsey, and S. Struben. Software tools for craniofacial growth and reconstruction. In *Graphics Interface*, june 1998. Vancouver, Canada. 26, 29, 32, 34, 35

[33] D.W. Bullock. *Computer Assisted 3D Craniofacial Reconstruction*. PhD thesis, The Univerity of British Columbia, October 1999. 26, 29, 32, 34

[34] K. kahler, J. Haber, and H.P. Seidel. Reanimating the dead: reconstruction of expressive faces from skull data. *ACM Transactions on Graphics*, 22(3):554–561, 2003. 26, 29, 34, 35

[35] Y. Pei, H. Zha, and Z. Yuan. Tissue map based craniofacial reconstruction and facial deformation using rbf network. In *Proc. of the third International Conference on Image and Graphics*. ICIG, 2004. 26, 29, 32, 34, 35

[36] S. Michael and M. Chen. The 3d reconstruction of facial features using volume distortion. In *Proc. 14th Eurographics UK Conference*, pages 297–305, 1996. 27, 28, 29, 32

[37] L.A. Nelson and S.D. Michael. The application of volume deformation to three-dimensional facial reconstruction : a comparison with previous techniques. *Forensic Science International*, 94:167–181, 1998. 27, 28, 29, 32, 34, 36

[38] G. Quatrehomme. A fully three-dimensional method for facial reconstruction based on deformable models. *Journal of Forensic Science*, 42(4):649–652, 1997. 27, 28, 29, 32, 34, 36

[39] G. Subsol and G. Quatrehomme. *Automatic 3D facial reconstruction by feature-based registration of a reference head, in: J.G Clement, M.K. Marks, Computer-Graphic Facial Reconstruction*, chapter 5, pages 79–101. Elsevier Academic Press, 2005. 27, 28, 29, 32, 34, 36, 37, 38

[40] G. Attardi, M. Berto, M. Forte, R. Gori, A. Guidazzoli, S.Imboden, and F. Mallegni. 3d facial reconstruction and visualization of ancient egyptian mummies using sprial ct data. In *Proc SIGGRAPH*, 1999. Los Angeles, USA. 27, 28, 29, 32, 34, 35, 36

[41] P. Vanezis, M. Vanezis, G. McCombe, and T. Niblet. Facial reconstruction using 3-d computer graphics. *Forensic Science International*, 108:81–95, 200. 27, 28, 29, 31, 34, 35

[42] M.W. Jones. Facial reconstruction using volumetric data. In *Proc. of the 6th International Vision Modeling and Visualisation Conference*, pages 21–23, November 2001. Stuttgart. 27, 28, 29, 32, 34, 36

[43] J. Muller, A. Mang, and T. M. Buzug. A template-deformation method for facial reproduction. In *Proc. of the 4th Int. Symposium on Image and Signal Processing and Analysis*, pages 359–364. ISPA, September 15-17 2005. Zagreb, Croatia. 27, 28, 32, 34, 35

[44] D. Vandermeulen, M. Loubele, P. Claes, Q. Wang, W. Mollemans, S. Srivastava, S. De Greef, G. Willems, and P. Suetens. Low-dose ct based soft tissue modeling for craniofacial reconstruction. In *2nd international conference on reconstruction of soft facial parts*. RSFP, March 15-18 2005. Remagen, Germany. 27, 29, 32, 35

[45] D. Vandermeulen, P. Claes, S. De Greef, G. Willems, and P. Suetens. Volumetric deformable face models for cranio-facial reconstruction. In *Proc. 4th international symposium on image and signal processing and analysis*, pages 353–358. ISPA, September 15-17 2005. Zagreb, Croatia. 27, 29, 32, 35, 36

[46] D. Vandermeulen, P. Claes, D. Loeckx, S. De Greef, G. Willems, and P. Suetens. Computerized craniofacial reconstruction using ct-derived implicit surface representations. *Forensic science international*, 159(1):164–174, 2006. 27, 29, 32, 35, 36

[47] W.D. Turner, R.E.B Brown, T.P. Kelliher, P.H. Tu, M.A. taister, and K.W.P Miller. A novel method of automated skull registration for forensic facial approximation. *Forensic Science International*, 154:149–158, 2005. 27, 32, 34, 36, 37

[48] P. Claes, D. Vandermeulen, P. Suetens, S. De Greef, and G. Willems. Computerized facial approximation using statistical models of tissue depth and 3-d facial outlook. In *Conference of the international association for craniofacial identification*, pages 91–96. IACI, October 2004. Dalian, China. 27, 29

[49] M. Berar, M. Desvignes, G. Bailly, and Y. Payan. 3d statistical facial reconstruction. In *Proceedings 4th international symposium on image and signal processing and analysis*. ISPA, 2005. Zagreb, Croatia. 27, 29, 32

[50] M. Berar, M. Desvignes, G. Bailly, and Y. Payan. 3d semi-landmark-based statistical face reconstruction. *Journal of computing and information technology*, 14(1):31–43, 2006. 27, 29, 32, 34, 36

[51] T. Hutton, B. Buxton, and P. Hammond. Estimating average growth trajectories in shape-space using kernel smoothing. *IEEE Transactions on Medical Imaging*, 22(6):747–753, 2003. 28

[52] A. J. Tyrell, M.P. Evison, A.T. Chamberlain, and M.A. Green. Forensic three-dimensional facial reconstruction : historical review and contemporary developments. *Journal of Forensic Science*, 42(4):653–661, 1997. 29

[53] P. Tu, R. Hartley, W. Lorensen, M. Allyassin, R. Gupta, and L. Heier. *Face Reconstruction using flesh deformation modes: in J.G Clement, M.K. Marks, Computer-Graphic Facial Reconstruction*, chapter 8, pages 145–162. Eslevier Academic Press, 2005. 29, 32, 34, 35

[54] A. Mang, J. Muller, and T.M. Buzug. A multi-modality computer-aided framework towards postmortem identification. *journal of computing and information technology*, 14(1):7–19, 2006. 29, 32, 34, 35, 37

[55] F. Bookstein. *Morphometric tools for landmark data*. Cambridge University Press, 1991. 32

[56] J.L. Plasencia. Landmark-based 3d mesh warping for bone-skin reconstruction. In *Proc. WSCG*, 1999. 32, 34

[57] M. Berar, M. Desvignes, G. Bailly, and Y. Payan. statistical skull models from 3d x-ray images. In *2nd international conference on reconstruction of soft facial parts*. RSFP, March 17-18 2005. Remagen, Germany. 32, 36

[58] M. Cairns. An investigation into the use of 3d computer graphics for forensic facial reconstruction. Technical report, Department of computer science, University of Glasgow, 1999. 34

[59] P.J. Besl and N.D. Mckay. A method of registration of 3-d shapes. *IEEE transactions on pattern analysis ans machine intelligence*, 14(2):239–256, February 1992. 36, 67, 74, 82, 238, 240

[60] W. Mollemans, F. Schutyser, N. Nadjmi, F. Maes, and P. Suetens. Soft tissue predictions with a tetrahedral mass tensor model for a maxillofacial planning system: a quantitative validation study. In *Proc. of SPIE*, february 2006. 39

[61] M. Chabanas, C. Marécaux, F. Chouly, F. Boutault, and Y. Payan. Evaluating soft tissue simulation in maxillofacial surgery using preoperative and postoperative ct scans. In *Proc. of CARS*, pages 419–424, 2004. 39

[62] S. Lele and J. Richtsmeier. *An invariant approach to statistical analysis of shapes*. Chapman&Hall / CRC, 2001. 39

[63] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 947–954, 2005. 39

[64] A. Hilton, A.J. Stoddart, J. Illingworth, and Windeatt. Marching triangles: range image fusion for complex object modelling. In *T. IEEE, International Conference on Image Processing*, 1996. 47

[65] J.C. Carr, R.K. Beatson, J.B. Cherrie, T.J. Mitchell, W.R. Fright, B.C. McCallum, and T.R. Evans. Reconsstruction and representation of 3d objects with radial basis functions. In *In computer graphics SIGGRAPH 2001, ACM Press/ACM SIGGRAPH*, pages 67–76, 12-17 August 2001. 48, 50

[66] R.K. Beatson, W.A. Light, and S. Billings. Fast solution of the radial basis function interpolation equations: Domain decomposition methods. *SIAM J.SCI.Comput.*, 22:1717–1740, 2000. 49

[67] R.K. Beatson and L. Greengard. *A short course on fast multipole methods, In M. Ainsworth, J. Levesley, W.A. Light, and M.Marletta: Wavelets, Multilevel Methods and Elliptic PDEs*, pages 1–37. 1997. Oxford University Press. 49

[68] http://www.farfieldtechnology.com/products/toolbox/. 49, 111

[69] J.C. Carr, R.K. Beatson, B.C. McCallum, W.R. Fright, T.J. McLennan, and T.J. Mitchell. Smooth surface reconstruction from noisy range data. In *In Pr. Computer Graphics SIGGRAPH 2003*, 2003. 50, 51, 75, 92, 104

[70] G. Turk and J.F. O'Brien. Variational implicit surfaces. Technical report, GIT-GVU-99-15, Georgia Institute of Technology, 1999. 50, 52

[71] H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 44–51. CVPR, 2000. 54, 108, 109, 110, 114, 249

[72] P.J. Huber. *Robust Statistics, in Wiley series in probabilityand mathematical statistics*. New York, 1981. 59

[73] R. Fransens, C. Strecha, and L. Van Gool. Robust estimation in the presence of spatially coherent outliers. In *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, 2006. 61, 221, 222, 223

[74] M.J. Black. *Robust incremental optical flow*. PhD thesis, Yale University, 1992. 65, 215, 217, 218

[75] M. J. Black and A. Rangarajan. The outlier process: Unifying line processes and robust stastistics. In *In Pr. CVPR 94*, 1994. 65, 215, 217, 218

[76] D. Geiger and A. Yuille. A comon framework for image segmentation. *International Journal of Computer Vision*, 6(3):227–243, 1991. 65, 215, 218, 223

[77] A. Blake and A. Zisserman. *Visual Reconstruction*. The MIT press, Cambridge, Massachusetts, 1987. 65, 218

[78] R.P. Brent. *Algorithms for minimization without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall, 1973. 66

[79] D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrf's: surface reconstruction. *Transactions on Pattern Analysis and Machine Intelligence*, 13(5):401–412, 1991. 68

[80] J. Vanden Wyngaerd and L. Van Gool. Automatic crude patch registration. *Computer Vision and Image Understanding*, 87(1-3):8–26, 2002. 74, 75, 239, 240, 242, 245

[81] D. Tubic, P. Hebert, and Denis Laurendeau. A volumetric approach for interactive 3d modeling. *Computer Vision and Image Understanding*, 92:56–77, 2003. 75, 86, 103, 238, 242, 243, 245

[82] T. Masuda. Registration and integration of multiple range images by matching signed distance fields. *Computer Vision and Image Understanding*, 87(1-3):51–65, 2002. 75, 86, 242, 243, 245

[83] G. Barequet and M. Sharir. Partial surface matching using directed footprints. In *In Pr. of the twelfth anual asymposium on computational geometry*, 1996. 75, 77, 80, 241, 245

[84] A.E. Johnson. *Spin-images: A representation for 3D matching*. PhD thesis, Robotics Institute Carnegie Mellon University, Pittsburgh, 1997. 75, 77, 78, 245

[85] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *In proceedings of the 23rd Annual Conference on Computer Graphics*, 1996. 88, 89, 90, 243, 244

[86] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *In Proceedings of SIGGAPH'94*, 1994. 89

[87] Bill Lorensen and H.E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *In Proceedings of ACM Siggraph*, 1987. 89, 243

[88] A. Hilton, A.J. Stoddart, J. Illingworth, and Windeatt. Marching traingles: Range image fusion for complex object modelling. In *IEEE International Conference on Image Processing*, 1996. 91

[89] Andrew Edie Johnson. *Spin-images: A representation for 3-D Surface Matching*. PhD thesis, Robotics Insititue, Carnegie Mellon Universaty, Pittsburgh, PA, 1997. 92, 241, 243, 244

[90] A. Johnson and S.B. Kang. Registration and integration of textured 3-d data. Technical report, CRL 96/4, Cambridge Research Laboratory, 1996. 92, 243, 244

[91] K. Pulli, T. Duchamp, H. Hoppe, J. McDonald, L. Shapiro, and W. Stuetzle. Robust meshes from multiple range maps. In *Proc. Int. Conf. on recent Advances in 3-D Digital Imaging and Modeling (3DIM)*, 1997. 92, 243, 244

[92] http://www.eyetronics.com. 94

[93] P. De Groeve, F. Schutyser, J. Van Cleynenbreugel, and P. Suetens. Registration of 3d photographs with spiral ct images for soft tissue simulation in maxillofacial surgery. In *In proceedings of MICCAI*, 2001. 95, 100

[94] A. W. F. Lee, D. Dobkin, W. Swelders, and P. Schroder. Multiresolution mesh morphing. In *In proc. SIGGRAPH*, 1999. 108, 248

[95] T. Hutton, B. Buxton, and P. Hammond. Dens surface point distribution models of human faces. In *IEEE workshop on mathematical methods in biomedical image analysis*, 2001. 108, 248

[96] C. Lorenz and N. Krahnstover. Generation of point-based 3d statistical shape models for anatomical objects. *Computer Vision and Image Understanding*, 77:175–191, 2000. 108, 248

[97] V. Blanz and T. Vetter. A morphable model for the syntesis of 3d faces. In *In Proc. SIGGRAPH*, 1999. 108, 248

[98] J. Paterson and A. Fitzgibbon. 3d head tracking using non-linear optimization. In *In Britisch Machine Vision Conference*, 2003. 108, 248

[99] S. Tondeur, P. Claes, D. Vandermeulen, and P. Suetens. Regularized non-rigid icp using variational implicit surfaces for model building. Technical report, KUL/ESAT/PSI/0601, January 2006. Leuven, Belgium. 108

[100] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. Technical report, Microsoft Research, Cambridge, 1999. 128

[101] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992. 129

[102] T. J. Hutton. *Dense surface models of the human face*. PhD thesis, Biomedical Informatics Unit. Eastman Dental Institute, University college London, 2004. 132

[103] J. Hug, C. Brechbhler, and G. Szkely. Tamed snake: A particle system for robust semi-automatic segmentation. In *Proc. of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 106–115. MICCAI, 1999. 141

[104] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003. 162

[105] Ralf W. Schmitz, Verlag Philipp von Zabern, and Mainz am Rhein. *Neanderthal 1056-2006*. ISBN-13:978-3-8053-3667-3. 178

[106] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8):677–688, 2001. 215, 225

[107] F.R. Hampel, E. M. Ronchetti, P.J. Rousseuw, and W. A. Stahel. *Robust Statistics: The approach based on influence functions*. John Wiley and Sons, New York, NY, 1986. 216

[108] G. Celeux, F. Forbes, and N. Peyrard. Em procedures using mean field-like approximations for markov model-based image segmentation. Technical report, RR-4105 Inria Rhone-Alpes, 2001. 223

[109] J. Zhang. The mean filed theory in em procedures for blind markov random field image restoration. *Signal Processing*, 40(10):2570–2583, 1992. 223

[110] R.E. Brown, T.P. Kelliher, P.H. Tu, W.D. Turner, M.A. Taister, and K.W.P. Miller. A survey of tissue-depth landmarks for facial approximation. . *Forensic Sci Communications*, 6(1), 2004. 230

[111] J. Kollmann and W. Buchly. *Die persistenz der rassen und die reconstruction der physiognomie prahistorischer schadel. Archiv fur anthropologie 1898;25:329-59. In: Taylor KT. Forensic art and illustration.*, chapter 25, pages 329–359. 2001. 230

[112] Yang Chen and Gerard Médioni. Object modeling by registration of multiple range images. In *In Proceedings of the 1991 IEEE International Conference on robotics and automatisation*, 1991. 238, 242

[113] E. Trucco, A.Fusiello, and V. Roberto. Robust motion and correspondence of noisy 3-d point sets with missing data. *Pattern Recognition Letters*, 20(9):899–898, 1999. 238

[114] Ko Nishino and Katsushi Ikeuchi. Robust simultaneous registration of multiple range images. In *ACCV2002, The 5th Asian Conference on Computer Vision*, 2002. 238

[115] Gerarld Dalley and Patrick Flynn. Pair-wise range image registration: A study in outlier classification. *Computer Vision and Image Understanding*, 87(1-3):104–115, 2002. 238

[116] Chitra Dorai, Gang Wang, Anil Jain, and Carolyn Mercer. Registration and integration of multiple object views for 3d model construction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):83–89, 1998. 238, 239

[117] Kari Pulli. Multi-view registration for large datasets. In *In Proceedings of the second International Conference on 3-D Imaging and Modelling, 3DIM'99*, 1999. 238, 239, 242

[118] Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *In Proceedings of ACM Siggraph 1994*, 1994. 238, 239

[119] Zhengyou Zhang. Iterative point matching for registration of free-form curves. Technical report, Technical Report RR-1658, INRIA, Epidaure, 1992. 238

[120] P.J. Neugebauer. Geometrical cloning of 3d objects via simultaneous registration of multiple range images. In *In Proceedings of the international Conference on Shape Modeling and Applications*, 1997. 238, 239

[121] Christian Schütz, T.Jost, , and Heinz Hügli. Semi-automatic 3d object digitizing system using range images. In *In Proc. ACCV'98*, 1998. 238

[122] editor Nicholas Ayache. *Artificial Vision for Mobile robots - Stero-vision and Multisensor Perception*. MIT Press, 1991. 239

[123] Yonghuai Liu and Marcos A. Rodriguez. Accurate registration of structerd data using two overlapping range images. In *In Proceedings of the 2002 IEEE International Conference on Robotics and automatation*, 2002. 239

[124] Yonghhuai Liu and Fred Labrosse. Inverse validation for accurate range image registration with structered data. In *In Proceedings of 16th International Conference on Pattern Recognition*, 2002. 239

[125] Thomas Pajdla and Luc Van Gool. Matching of 3-d curves using semi-differential invariants. In *In Proceedings of the fifth International Conference on Computer Vision*, 1995. 239

[126] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *In proceedings of the third International Conference on 3-D Digital Imaging and Modeling, 3DIM*, 2001. 239

[127] Christian Shütz and T.Jost. Multi-feature matching algorithm for free-form 3d surface registration. In *In Proceedings of the 14th International Conference on Pattern Recognition*, 1998. 239

[128] David W. Eggert, Andrew W. Fitzgibbon, and Robert B.Fisher. Simultanous registration of multiple range views for use in reverse engineering of cad models. *Computer Vision and Image Understanding*, 69(3):253–272, 1998. 239, 242

[129] Takeshi Masuda and Naokazu Yokoya. A robust method for registration and segmentation of multiple range images. *Computer Vision and Image Understanding*, 61(3):295–307, 1995. 239

[130] Guy Godin, Denis Laurendau, and Robert Bergevin. A method for the registration of attributed range images. In *In Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling (3DIM 2001)*, 2001. 239

[131] Gérard Blais and Martine Levine. Registering multiview range data to create 3d computer objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):820–824, 1995. 239, 242

[132] Raouf Benjemaa and Francis Schmitt. Fast global registration of 3d sampled surfaces using a multi-z-buffer technique. *Image and Vision Computing*, 17:113–123, 1999. 239, 242

[133] P. Voila and W. Wells. Alignment by maximisation of mutual information. In *In Proceedings of the International Conference on Computer Vision*, 1999. 239

[134] R.J. Campbell and Patrick Flynn. A survey of free-form object representation and recognition techniques. *Computer Vision and Image Understanding*, 81(2):166–210, 2001. 239

[135] A. Johnson and Hébert. Surface matching for object recognition in cluttered scenes. *IEEE Transaction on Pattern Analysis and Machine intelligence*, 10(5), 1999. 239, 240, 241

[136] M. Wheeler. *Automatic Modeling and Localization for Object Recognition*. PhD thesis, School of Computer Science, Carnegie Mellon UNiversaty, 1996. 239

[137] M. Garland and P. Heckbert. Surface simplification using quadric error metrics. In *In Proceedings of Siggraph 97*, 1997. 239

[138] D.F. Huber and M. Hebert. A new approach to a 3-d terrain mapping. In *In Proceedings of the 1999 IEEE/RSJ International Conference on intelligent Robotics and Systems*, 1999. 239

[139] Gérard Blais and Martine Levine. Registering multiview range data to create 3d computer objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):820–824, 1995. 239

[140] Craig Robertson and Robert B. Fisher. Parallel evolutionary registration from range data. *Computer Vision and Image Understanding*, 87(1-3):39–50, 2002. 240

[141] Chu-Song Chen, Yi-Ping Hung, and Jen-Bo Cheng. Ransac-based darces: A new approach to fast automatic registration of partially-overlapping range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1229–1234, 1999. 240

[142] L. Lucchese, G. Doretto, and G.M. Cortelazzo. Frequency domain estimation of 3-d rigid motion based on range and intensity data. In *In Proceedings of the international Conference of Recent Advances in 3-D Digital Imaging and Modeling (3DIM'97)*, 1997. 240

[143] S.B. Kang and K. Ikeuchi. 3-d object pose determination using complex egi. Technical report, CMU-RI-TR-90-18, Robotics Institute, CarnegieMellon University, 1990. 240

[144] K. Ikeuchi. Recognition of 3-d objects using the extended gaussian image. In *In Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1997. 240

[145] J.p. Thirion. Extremal points:definition and application to 3d image registration. In *Proceedings of IEEE conference on computer vision and pattern recognition*, 1994. 240

[146] Jacques Feldmar, Nicholas Ayache, and Fabienne Betting. 3d-2d projective registrationof free formcurves and surfaces. Technical report, No. 2434, INRIA, Epidaure, 1994. 240

[147] Pavel Kresk, Thomas Padja, and Vaclav Hlavac. Differential invariants as the base of triangulated surface registration. *Computer Vision and Image Understanding*, 87(1-3):27–38, 2002. 240

[148] C.S.Chua and R. Jarvis. Point-signatures: a new representation for 3d object recognition. *International Journal of Computer Vision*, 25(1):63–85, 1997. 240

[149] Y.Sun and M.A.Abidi. Surface matching by 3d point's fingerprint. 240

[150] A.E. Johnson and M. Hébert. Surface registration by matching oriented points. In *In International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, 1997. 240, 241

[151] A.E. Johnson and M. Hébert. Recognizing objects by matching orieted points. In *In Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1997. 240, 241

[152] S.M. Yamany and A.A. Farag. Free-from surface registration using surface signatures. In *IEEE International Conference on Computer Vision*, 1999. 241

[153] Sameh M. Yamany and Aly A. Farag. Surface signatures: An orientation independent free-form surface repressentation scheme for the purpose of objects registration and matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1105–1120, 2002. 241

[154] A.P.Ashbrook, Robert B. Fisher, Craig Robertson, and N.Werghi. Finding surface correspoondences for object recognition and registration using pairwise geometric histograms. In *In Proceedings of the european Confernece on Computer Vision (ECCV'98)*, 1998. 241

[155] Robert Bergevin, Denis Laurendau, and D.Poussart. Registering range views of multipart objects. *Computer Vision and Image Understanding*, 61(1):1–6, 1995. 241

[156] Pvel Kresk, Thomas Padja, Vaclav Hlavac, and Ralph Martin. Range image registration driven by hierarchy of surface differential features. In *In Proceedings of the 22nd Workshop of the Austrain Association for Pattern Recognition*, 1998. 242

[157] Daniel Hubert and Martial Hébert. Fully automatic registration of multiple 3d data sets. In *In IEEE Computer Society Workshop on Computer Vision Beyond The Visible Spectrum*, 2001. 242

[158] Robert Bergevin, Marc Soucy, H. Gagnon, and Denis Laurendeau. Towards a general multiview registration technique. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 18(5):540–547, 1996. 242

[159] H. Gagnon, Marc Soucy, Robert Bergevin, and Denis Laurendeau. Registration of multiple range views for automatic 3-d modelbuilding. In *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, 1994. 242

[160] H.Shum, Martial Hébert, Katsushi Ikeuchi, and R. Reddy. An integral approach to free-form object modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1366–1370, 1997. 242

[161] Andrew Edie Johnson and Sing Bing Kang. Registration and integration of textured 3-d data. *Image and Vision Computing*, 17(2), 1999. 242

[162] O. Jokinen. Area-based matching for simultaneous registration of multiple 3-d profile maps. *Computer Vision and Image Understanding*, 71(3):431–447, 1998. 242

[163] P.J. Neugebauer. Reconstruction of real-world objects via simultaneous registration and robust combination of multiple range images. *International Journal Shape Modeling*, 3 (1-2):71–90, 1997. 242

[164] J. Williams and M.Bennamoun. Simultaneous registration of multiple corresponding point sets. *Computer Vision and Image Understanding*, 81(1):117–142, 2001. 242

[165] J.D. Boissonnat. Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics*, 3(4):266–286, 1984. 242

[166] R. Veltcamp. 2d and 3d object reconstruction with the γ-neighborhood graph. Technical report, CS-R9116, CWI Centre for Mathematics and Computer Sience, 1991. 242

[167] Hugues Hoppe. *Surface reconstruction from unorganized points*. PhD thesis, Universaty of Washington, Department of Computer Sience and Engineering, 1994. 242

[168] C.L. Bajaj, F. Bernardini, and G. Xu. Automatic reconstruction of surfaces and scalar fields from 3d scans. In *In proceedings of SIGGRAPH'95*, 1995. 242

[169] M. Soucy and D. Laurendeau. A general surface approach to the integration of a set of range views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4):344–358, 1995. 243

[170] M. Rutishauser, M. Stricker, and M. Trobina. Merging range images of arbitrarily shaped objects. In *Proc. IEEE Computer Vision and Pattern Recognition(CVPR'94)*, 1994. 243

[171] A. Hilton, A. Stodddart, J. Illingworth, and T. Windeatt. Reliable surface reconstruction from multiple range images. In *In pr. Computer Vision -ECCV '96*, 1996. 243

[172] Wheeler. Object shape and reflectance modeling from observation. In *Kluwer Engineering And Computer Science International Series*, 2001. 243

[173] Takashi Masuda. Filling the signed distance field by fitting local quadrics. In *In proceedings of the 2nd international symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04)*, 2004. 244

[174] Emiliano D'Agostino, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. An information theoretic approach for non-rigid image registration using voxel class probabilities. *Med Image Anal*, xx:xxxx, 2005. 247

[175] Dirk Loeckx, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Non-rigid image registration using free-form deformations with a local rigidity constraint. volume 3216, pages 639–646, 2004. 247

[176] R. Davies, C. Twinning, T. Cootes, J. Waterton, and C. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, 2002. 248

[177] Y. Wang, B. Peterson, and L. Staib. Shape-based 3d surface correspondence using geodesics and local geometry. *Computer Vision and Pattern Recognition*, 2:644–651, 2000. 248

[178] A. Brett, A. Hill, and C. Taylor. A method of 3d surface correspondence for automated landmark generation. In *In British Machine Vision Conference*, 1997. 248

[179] A. Brett and C. Taylor. A method of autromated landmark generation for automated 3d pdm construction. In *In British Machine Vision Conference*, 1998. 248

[180] E. Huot, H. Yahia, I. Cohen, and I. Herlin. Matching structures by computing minimal paths on a manifold. Technical report, INRIA. 248

[181] J.A.K. Suykens and J. Vandewalle. Least square support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999. 251

[182] http://www.esat.kuleuven.ac.be/sista/lssvmlab/. 252

Publications

International Journal

1. S. De Greef, P. Claes, W. Mollemans, M. Loubele, D. Vandermeulen, P. Suetens, G. Willems, Semi-automated ultrasound facial soft tissue depth registration : method and validation, *Journal of forensic sciences*, vol. 50(6), 7pages, 2005

2. S. De Greef, P. Claes, D. Vandermeulen, W. Mollemans, P. Suetens, G. Willems, Large-scale in-vivo caucasian facial soft tissue thickness database for craniofacial reconstruction, *Forensic science international*, vol. 159 supplement 1, pp. S126-S146, 2006

3. D. Vandermeulen, P. Claes, D. Loeckx, S. De Greef, G. Willems and P. Suetens. Computerized craniofacial reconstruction using CT-derived implicit surface representations, *Forensic Science International*, vol. 159 supplement 1, pp. S164-S174. 2006

4. P. Claes, D. Vandermeulen, S. De Greef, G. Willems, P. Suetens, Statistically deformable face models for cranio-facial reconstruction, *Journal of computing and information technology*, vol. 14, no. 1, pp. 21-30, 2006

5. P. Claes, D. Vandermeulen, S. De Greef, G. Willems, P. Suetens, Craniofacial reconstruction using a combined statistical model of face shape and soft tissue-depths : methodology and validation, *Forensic science international*, vol. 159 supplement 1, pp. S147-S158, 2006

International Journal Abstract

1. S. De Greef, P. Claes, W. Mollemans, D. Vandermeulen, P. Suetens, G. Willems, In vivo semi-automated ultrasound facial soft tissue depth registration : method and preliminary results, *Journal of Anatomy* - proceedings of the Anatomical Society of Great-Britain and Ireland , vol. 5, no. 206, pp. 510, 2005 (Winter meeting of the Anatomical Society of Great-Britain and Ireland, January 5-7, 2005, Oxford, UK)

National Journal

1. S. De Greef, P. Claes, W. Mollemans, D. Vandermeulen, P. Suetens, G. Willems, Computergestuurde forensische aangezichtsreconstructie : recente ontwikkelingen en tendensen, *Belgisch tijdschrift voor tandheelkunde*, jaargang 60/3, pp. 237-249, 2005, VVT Brussel

2. S. De Greef, P. Claes, W. Mollemans, D. Vandermeulen, P. Suetens, G. Willems, Reconstitutions faciales assistes par ordinateur : dveloppements rcents et tendances, *Revue Belge de mdecine dentaire*, vol. 60/3, 2005, Socit de mdecine dentaire

International Conference : Published in Proceedings

1. P. Claes, S. Degreef, G. Willems, D. Vandermeulen, P. Suetens, Craniofacial statistical modeling and reconstruction, *3D modelling*, 11 pp., April 28-29, 2004, Paris, France

2. P. Claes, D. Vandermeulen, P. Suetens, S. De Greef, G. Willems, Computerized facial approximation using statistical models of tissue depth and 3-D facial outlook, *Conference of the international association for craniofacial identification* - IACI'2004, pp. 91-96, October 2004, Dailan, China

3. D. Vandermeulen, M. Loubele, P. Claes, Q. Wang, W. Mollemans, S. Srivastava, S. De Greef, G. Willems, P. Suetens, Low-dose CT based soft tissue modeling for craniofacial reconstruction, *2nd international conference on reconstruction of soft facial parts* - RSFP 2005, March 17-18, 2005, Remagen, Germany

4. P. Claes, D. Vandermeulen, S. De Greef, G. Willems, P. Suetens, Combined statistical modeling of tissue depth and 3D facial outlook for computerized facial approximation, *2nd international conference on reconstruction of soft facial parts* - RSFP 2005, March 17-18, 2005, Remagen, Germany

5. S. De Greef, P. Claes, W. Mollemans, D. Vandermeulen, P. Suetens, G. Willems, Validated semi-automated ultrasound facial soft tissue depth registration, *2nd international conference on reconstruction of soft facial parts* - RSFP 2005, March 17-18, 2005, Remagen, Germany

6. P. Claes, D. Vandermeulen, L. Van Gool, P. Suetens, Partial surface integration based on variational implicit functions and surfaces for 3D model building, *Proceedings 5th international conference on 3-D digital imaging and modeling* - 3DIM 2005, pp. 31-38, June 13-17, 2005, Ottawa, Ontario, Canada

7. P. Claes, D. Vandermeulen, L. Van Gool, P. Suetens, Robust and accurate partial surface registration based on variational implicit surfaces for automatic 3D model building, *Proceedings 5th international conference on 3-D digital imaging and modeling* - 3DIM 2005, pp. 385-392, June 13-17, 2005, Ottawa, Ontario, Canada

8. D. Vandermeulen, P. Claes, P. Suetens, S. De Greef, G. Willems, Volumetric deformable face models for cranio-facial reconstruction, *Proceedings 4th international symposium on image and signal processing and analysis* - ISPA2005, pp. 353-358, September 15-17, 2005, Zagreb, Croatia

9. P. Claes, D. Vandermeulen, S. De Greef, G. Willems, P. Suetens, Statistically deformable face models for cranio-facial reconstruction, *Proceedings 4th international symposium on image and signal processing and analysis* - ISPA2005, pp. 347-352, September 15-17, 2005, Zagreb, Croatia

10. J. Hermans, P. Claes, J. Bellemans, D. Vandermeulen, P. Suetens, Robust initialization for 2D/3D registration of knee implant models to single-plane fluoroscopy, *SPIE medical imaging 2007 conference*, February 17-22, 2007, San Diego, California, USA

11. J. Hermans, P. Claes, J. Bellemans, D. Vandermeulen, P. Suetens, A robust optimization strategy for intensity-based 2D/3D registration of knee implant models to single-plane fluoroscopy, *SPIE medical imaging 2007 conference*, February 17-22, 2007, San Diego, California, USA

International Conference : Abstract or Not Published

1. S. De Greef, P. Claes, D. Vandermeulen, P. Suetens, G. Willems, Semi-automated ultrasound facial soft tissue depth registration, *8th Cross Channel Conference - CCC2004, international conference on legal medicine and forensic sciences*, April 20-24, 2004, Brugge, Belgium

2. S. De Greef, P. Claes, D. Vandermeulen, P. Suetens, G. Willems, Semi-automated ultrasound facial soft tissue depth registration : method and preliminary results, *Conference of the international association for craniofacial identification* - IACI'2004, October 2004, Dailan, China

3. S. De Greef, P. Claes, W. Mollemans, D. Vandermeulen, P. Suetens, G. Willems, Semi-automated ultrasound facial soft tissue depth registration : method and preliminary results, *57th annual meeting of the American Academy of Forensic Sciences*, February 21-26, 2005, New Orleans, LA, USA

4. P. Claes, D. Vandermeulen, S. De Greef, G. Willems, P. Suetens, PCA-based statistical model for facial soft tissue depth database exploration and specification, *17th meeting of the International Association of Forensic Sciences*, IAFS2005, pp. 111 (ref A0772), August 21-26, 2005, Hong Kong, China

5. S. De Greef, D. Vandermeulen, P. Claes, W. Mollemans, G. Willems, A new facial soft tissue depth dataset for the Caucasian model, *17th meeting of the International Association of Forensic Sciences*, IAFS2005, pp. 110 (ref A0740), August 21-26, 2005, Hong Kong, China

6. D. Vandermeulen, S. De Greef, P. Claes, P. Suetens, G. Willems, Automated craniofacial reconstruction using ultrasound and CT-based soft tissue thickness databases, *12th biennial meeting of the international association of craniofacial identification* - IACI'2006, November 1-4, 2006, Turkey, Istanbul

National Conference : Abstract or Not Published

1. S. De Greef, P. Claes, W. Mollemans, D. Vandermeulen, P. Suetens, G. Willems, Semi-automatische ultrasoon gelaats zachte weefseldikten-registratie : methode en validatie, *Vijfde VVT (Verbond van Vlaamse Tandartsen) congres*, May 26-28, 2005, Oostende, Belgium

Internal Report

1. P. Claes, D. Vandermeulen, L. Van Gool, P. Suetens, Automatic, robust and accurate 3D modelling based on variational implicit surface, *Internal report KUL/ESAT/PSI/0405*, K.U.Leuven, ESAT, December 2004, Leuven, Belgium

2. S. Tondeur, P. Claes, D. Vandermeulen, P. Suetens, Regularized non-rigid ICP using variational implicit surfaces for model building, *Internal report KUL/ESAT/PSI/0601*, K.U.Leuven, ESAT, January 2006, Leuven, Belgium

3. P.Claes, D. Vandermeulen, R. Fransens, S. De Greef, G. Willems, L.Van Gool and P. Suetens, Robust craniofacial reconstruction using a statistical model embedded in an EM framework applied to a craniometric and an implicit skull representation, *Internal report KUL/ESAT/PSI/0702*, K.U.Leuven, ESAT, January 2007, Leuven, Belgium

Curriculum Vitae

Peter Claes was born on October 30, 1979. From 1990-1997, he studied mathematics and science at the Onze-Lieve-Vrouw college (OLVE) in Edegem. He started university in 1997 as an engineering student at the Katholieke Universiteit Leuven, Belgium, graduating in 2002 with distinction. He majored in electronic engineering, multimedia and signal processing. During his studies he followed a one-week intensive course on image processing part of the ATHENS program. He also accomplished three projects as a job student concerning software development at Siemens Mobile in Herentals (Belgium).

In October 2002, Peter Claes started working as a PhD student at the Medical Image Computing laboratory of the Katholieke Universiteit Leuven. Although the laboratory is part of the faculty of engineering, it is located in the premises of the University Hospital Gasthuisberg. The main application of his research is craniofacial modeling and reconstruction. The main topics of his research concern, robust surface registration and integration using implicit functions, statistical modeling, analysis and manipulation. The coming years he will work on a statistical facial morphology approach for proof of identity across an ethnically diverse population at the Dental Science department of the University of Melbourne, Australia. His main domain of application for the future is Forensic Science and Identification.

Peter Claes is author or co-author of 5 international journal papers, 2 national journal papers and 11 papers in international peer-reviewed conference proceedings, while he also has contributions to several other national and international conferences. He has attended the 3D modeling conference in 2004, the IACI (International association for craniofacial identification) conference in 2004, the RSFP (international conference on reconstruction of soft facial parts) conference in 2005 and 2006, the 3DIM (international conference on 3-D digital imaging and modeling) conference in 2005 and the ISPA (international symposium on image and signal processing and analysis) symposium in 2005.

Peter Claes' mother tongue is Dutch, while he is also fluent in English and French. He has a basic knowledge of German. In his spare time, Peter Claes is highly passionate about mountain biking and being with friends.

Appendices

Appendix A

Robust Statistics for surface registration

In this appendix we present two main approaches for robust statistics and apply it to the surface registration problem. The first approach makes use of robust estimators that are substituted for the classical estimators such as the sum of squared errors (when the residuals are normally distributed). The second approach models the outliers explicitly as well, which allows the modeling of additional constraints such as, e.g. spatial coherence of outliers. Furthermore, it can be shown that for every such outlier process, an equivalent robust estimator from the first approach can be constructed.

A.1 Robust statistics using estimators

Robust statistics addresses parameter estimation problems using an estimator $\rho(x, \gamma)$ which is a function of residual errors x and a scale parameter γ. Robust statistics is widely applied for regression problems, however it is suitable to apply to different estimation problems as well, e.g optical flow [74], image reconstruction [75], image segmentation [76, 106], ... We will apply robust statistics to deal with outliers during surface registration. Setting the residual errors x equal to the measured point similarities $x = \mathfrak{s}_j(\boldsymbol{\theta})$, the robust statistics formulation of the surface registration problem amounts to:

$$\text{argmin}_{\boldsymbol{\theta}} \sum_{j=1}^{N} \rho(\mathfrak{s}_j(\boldsymbol{\theta}), \gamma) \tag{A.1}$$

When the residuals are normally distributed, the scale parameter is set equal to the noise standard deviation $\gamma = \sigma$ and the estimator becomes the optimal Quadratic estimator:

$$\rho(\mathfrak{s}_j(\boldsymbol{\theta}), \sigma) = \frac{\mathfrak{s}_j(\boldsymbol{\theta})^2}{2\sigma^2} \tag{A.2}$$

which gives rise to the standard Least-Squares estimation problem. The function ρ is called an M-estimator since it corresponds to the Maximum-Likelihood parameter estimate similar to our MAP parameter estimate in (3.23). Comparing (A.1) and (A.2) with the MAP objective function in (3.23) further similarities between both registration problem formulations are observed. However, two major differences are apparent. Firstly, as a result of the probabilistic prior transformation model approach, the MAP objective function in (3.23) is in fact a regularized instead of a standard least-square estimation. Secondly, as a result of the probabilistic data-likelihood approach, the scale parameter σ is treated as an extra ML-parameter to estimate, hence the $\log\sigma$ term in (3.23). In contrast, M-estimators typically (not always) use a multiple of the Median of Absolute Deviations (MAD) as a scale estimate which implicitly assumes that the noise contamination rate is 50 %. Due to the strong similarities between (3.23) and (A.2), outlier influence in (3.23) can be obtained by analyzing the outlier influence of the equivalent M-estimator (A.2). The latter is done according to the approach in [107], based on influence functions. The influence function characterizes the bias that a particular measurement has on the solution and is proportional to the derivative Ψ of the estimator ρ. For the Quadratic (Q) estimator $\rho(x,\sigma) = (\frac{x}{\sqrt{2}\sigma})^2$ this is:

$$\Psi(x,\sigma) = \frac{x}{\sigma^2} \tag{A.3}$$

which is illustrated in figure A.1(a) and (b). For least-squares estimation, regularized or not, the influence of outliers increases linearly and without bound (figure A.1(b)). To increase robustness against outliers, the estimator must be more insensitive to outlier measurements. A robust statistics approach is to select a robust M-estimator from a predefined cookbook of estimators from which the redescending M-estimators are a popular choice [107]. Two such redescending estimators examples are the Lorentzian:

$$\rho(x,\sigma) = \log(1 + \tfrac{1}{2}(\tfrac{x}{\sigma})^2) \quad \Psi(x,\sigma) = \tfrac{2x}{2\sigma^2 + x^2} \tag{A.4}$$

and the Truncated Quadratic (TQ):

$$\begin{aligned} \rho(x,\alpha,\beta) &= \begin{cases} \alpha x^2 & \text{if} \quad |x| < \sqrt{\beta}/\sqrt{\alpha} \\ \beta & \text{if} \quad |x| \geq \sqrt{\beta}/\sqrt{\alpha} \end{cases} \\ \Psi(x,\alpha,\beta) &= \begin{cases} 2\alpha x & \text{if} \quad |x| < \sqrt{\beta}/\sqrt{\alpha} \\ 0 & \text{if} \quad |x| \geq \sqrt{\beta}/\sqrt{\alpha} \end{cases} \end{aligned} \tag{A.5}$$

which are both illustrated in figure A.1(c,d) and (e,f) respectively. Observing the corresponding influence functions we see that both the estimators have a saturating property. The influence of outliers redescends to zero.

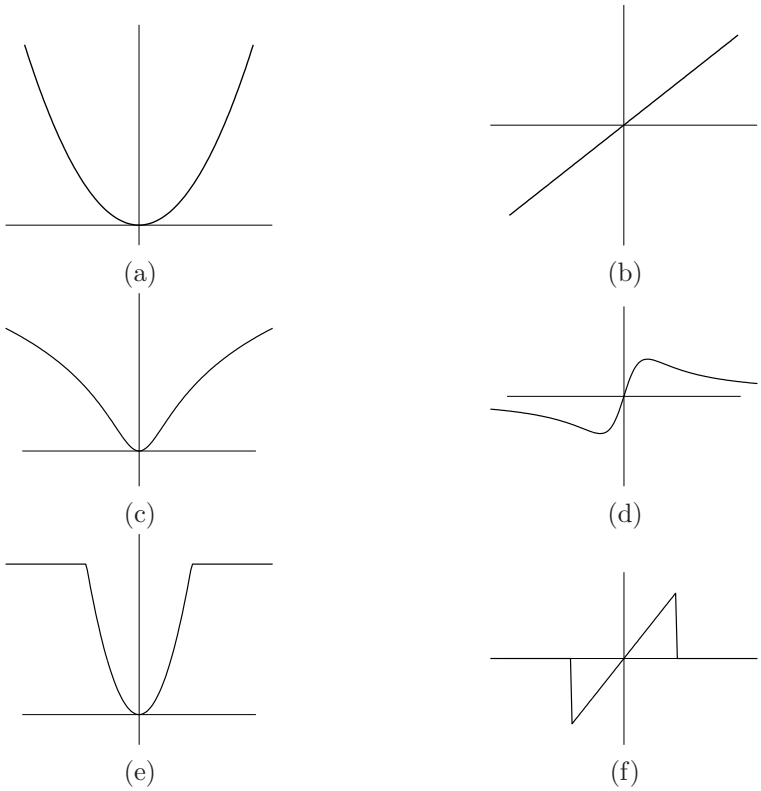


Figure A.1: Different M-estimators (left side) and their outlier influence functions (right side) (a,b) The Quadratic M-estimator with $\sigma = 1$, (c,d) The Lorentzian estimator with $\sigma = 1$ and (e,f) The Truncated Quadratic with $\alpha = 0.5$ and $\beta = 5$. The original Quadratic has a linear increasing outlier influence and without bound, while both the Lorentzian and the Truncated Quadratic have interesting saturating outlier influence properties.

A.2 Robust statistics using outlier-process modeling

Alternatively, instead of selecting a predefined estimator one could also model the outlier-process explicitly. According to [74, 75], a problem formulated in terms of explicit outlier processes can be converted or viewed in terms of robust statistics. An explicit outlier process formulation, however, is more general than the original robust estimation. For example, due to the explicit nature of the outlier process, constraints on the spatial organization of the outliers can be formulated. Due to the unification with robust estimation defined in [74, 75], equivalent M-estimators can be determined to analyze the influence of outliers within an outlier process. Furthermore, well known deterministic

continuation optimization routines for robust estimators (eg. deterministic annealing) can be directly applied to the explicit outlier process formulation.

An outlier process explicitly introduces an outlier latent variable z_j, which signals whether the residual x_j is an outlier, $z_j = 0$, or not, $z_j = 1$. This latent variable can either be binary, $z_j = \{0, 1\}$, suggesting a binary outlier process or continuous, $z_j = [0, 1]$, suggesting a continuous outlier process. A general outlier process formulation for transformation parameter $\boldsymbol{\theta}$ estimation is then defined as:

$$\text{argmin}_{(\theta,\mathbf{z})} \sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j)) \tag{A.6}$$

where

$$h(x, \alpha(z), \beta(z)) = \alpha(z)x^2 + \beta(z) \tag{A.7}$$

Depending on the definition of $\alpha(z)$ and $\beta(z)$ the outlier-process is either continuous or binary. An example for the latter is $\alpha(z) = z\alpha^*$ and $\beta(z) = (1 - z)\beta^*$. The outlier-process formulation in (A.6) leads to a joint estimation problem where besides the parameters $\boldsymbol{\theta}$ also the outlier map of latent variables $\mathbf{z} = \{z_j | j = 1, \ldots, N\}$ is to be estimated. According to [74, 75] and [77], the latent variables $\mathbf{z}$ can be removed from the equation by first minimizing over them, such that a new objective function solely in function of $\boldsymbol{\theta}$ is obtained. Alternatively, according to [76] the same result can be obtained by integrating the latent variables $\mathbf{z}$ out of the equation following a mean-field technique:

$$\begin{array}{l} \text{argmin}_{\boldsymbol{\theta}} \left(\text{argmin}_{\mathbf{z}} \sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j)) \right) \\ \text{argmin}_{\boldsymbol{\theta}} \left(\sum_{\mathbf{z}} \sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j)) \right) \end{array} \tag{A.8}$$

Doing so, leads to a dual-update of the surface registration problem. The update of the outlier map $\mathbf{z}$ is alternated with the update of the parameters $\boldsymbol{\theta}$ and during either update the other estimate is kept fixed. Furthermore, after minimization over $\mathbf{z}$, for example, an equivalent M-estimator can be defined as the infimum for various values z of a function $h(x)$:

$$\rho(x, \alpha, \beta) = \inf_{0 \leq z \leq 1} h(x, \alpha(z), \beta(z)) \tag{A.9}$$

The unified robust statistics formulation for the outlier process after latent variable elimination then becomes:

$$\text{argmin}_{\boldsymbol{\theta}} \sum_{j=1}^{N} \rho(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha, \beta) \tag{A.10}$$

Doing so, an outlier influence analysis can be performed. In figure A.2 the function $h(x, \alpha(z), \beta(z))$ with $\alpha(z) = z\alpha^*$ and $\beta(z) = (1 - z)\beta^*$ is plotted (dotted lines) for various values of the latent variable z and the infimum of

this family is plotted in bold. Note that the infimum can be determined solely based on the cases of h where $z = 1$ and $z = 0$, suggesting a binary valued outlier-process $z = \{0, 1\}$. It is clearly observed that the infimum strongly resembles the Truncated Quadratic (TQ) depicted in figure A.1(e) and defined in (A.5) with $\alpha = \alpha^*$ and $\beta = \beta^*$. Therefore, the defined binary outlier process shares the same outlier influence shown in figure A.1(f).

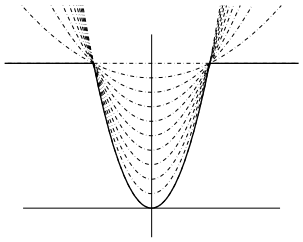


Figure A.2: Equivalent M-estimator for outlier processes obtained by minimizing over the outlier map $\mathbf{z}$. The bold line is the infimum of a function $h(x)$, while the dotted lines represent the evaluations of $h(x)$ with different values for the latent variable z. The parameters $\alpha = 0.5$ and $\beta = 5$ were set according to the same parameters of the Truncated Quadratic in figure A.1(e) and defined in (A.5).

Note that in robust statistics alternative estimators besides M-estimators exist. However due to the Maximum Likelihood nature of M-estimators, they are the most equivalent estimators to our MAP objective function formulation. A complete overview of robust statistics is outside the scope of this thesis. In the next section we will continue the MAP objective function formulation, while making use of explicit probabilistic outlier-processes and their equivalent M-estimators to handle outliers.

Appendix B

Outlier robust MAP estimation

In this appendix, following Fransens *et al.* [73], we will present a principled approach to link the robust estimators presented in appendix A to a statistical model of how the measurements, including the outliers, were generated. Indeed, the outlier-process will be modeled as a statistical process as well, akin to the inlier-process. As a result, making use of the theory expanded in appendix A, we will be able to construct equivalent M-estimators for outlier rejection. The final result will be an objective function with quasi self-regulating outlier robustness, depending on the kind of outliers, besides self-regulating noise robustness obtained previously.

B.1 Complete process

We will model both the inliers and outliers as random variables within the complete data log-likelihood resulting in a robust copy of the first term in (3.23). The transformation model-likelihood or the second term in (3.23), remains unchanged. Therefore, we will concentrate on the complete data log-likelihood term from now on.

Suppose a data sample $\mathfrak{s}_j$ is generated by either an inlier-process or an outlier-process, then the generating complete-process can be specified by conditioning the sample likelihood on the state of a binary-valued latent variable z_j, signalling whether a data sample $\mathfrak{s}_j$ was generated by the inlier-process $z_j = 1$ or not $z_j = 0$:

$$\begin{aligned} p(\mathfrak{s}_j|z_j,\Theta) &= \begin{cases} p(\mathfrak{s}_j|\Theta) & \text{if} \quad z_j = 1 \\ p^o(\mathfrak{s}_j) & \text{if} \quad z_j = 0 \end{cases} \\ p(\mathfrak{s}_j|z_j,\Theta) &= p(\mathfrak{s}_j|\Theta)^{z_j} p^o(\mathfrak{s}_j)^{(1-z_j)} \text{ for } z_j \in \{0,1\} \end{aligned} \tag{B.1}$$

$p(\mathfrak{s}_j|\boldsymbol{\theta})$ is a Gaussian[1] probability density function (PDF) constituting the inlier-process and was defined in (3.22). $p^o(\mathfrak{s}_j)$ is the outlier PDF constituting the outlier-process for which several choices can be considered. First off all, if some knowledge about the outlier generating process is known, then $p^o(\mathfrak{s}_j)$ can be set to a known distribution. In most cases, however, no outlier generating process knowledge is given and $p^o(\mathfrak{s}_j)$ can either be estimated similar to the inlier-distribution or can be set to a fixed uniform distribution for example. The latter is the case for our surface registration problems. Therefore, we take a fixed outlier distribution, which is assumed to be uniformly distributed $p^o(\mathfrak{s}_j) = \delta$. Spatial constraints on the outliers are typically incorporated using a proper modeling of z_j depending on the prior knowledge of the outlier problem besides the outlier PDF.

The latent variable (LV) can be non-probabilistic (deterministic) or probabilistic (random). Deterministic latent variables are typically used within robust statistics, while random latent variables are a theoretical continuation of the probabilistic objective function formulation. We will first present the derivation for random latent variables. Next, we discuss the special case of deterministic latent variables (dependent on the parameters $\boldsymbol{\theta}$ by a non-random function). Deterministic and random latent variable can then be combined into a single, augmented, outlier latent variable. We also show how the robust estimators presented in appendix A can be obtained starting from the latent variable modeling approach.

B.2 Random Bernoulli distributed LV

The most general type of latent variable we consider is a random (probabilistic) variable, where the binary latent variable z_j is Bernoulli distributed and the outliers are uniformly distributed ($p^o(\mathfrak{s}_j) = \delta$). The consequence is that the outlier map $\mathbf{z}$ is a random map with an associated prior-distribution $p(\mathbf{z})$. Let P be the prior probability of z_j being an inlier (i.e the fraction of observations $\mathfrak{s}_j$ thought to be generated by the inlier-process) and let $P_o = 1 - P$ be the prior probability of being an outlier (i.e the fraction of observations $\mathfrak{s}_j$ thought to be generated by the outlier-process). Then, assuming $\mathbf{z}$ is i.i.d[2], $p(\mathbf{z})$ can be specified as follows:

$$p(\mathbf{z}) = \prod_j^N P^{z_j} P_o^{(1-z_j)} \tag{B.2}$$

Note that spatial constraints on the outliers can be incorporated by choosing an alternative and appropriate prior-distribution. For example, in [73], $\mathbf{z}$ is considered a probabilistic random variable map modeled as a binary Markov Random Field with associated Gibbs-prior distribution ensuring spatially co-

[1] In general, $p(\mathfrak{s}_j|\boldsymbol{\theta})$ can be any probability density function, but in our application, we have assumed the measurement noise to be Gaussian, hence the choice of this particular PDF.

[2] i.i.d. = independent and identically distributed

herent outliers. Due to the probabilistic nature of the latent random variable we can write:

$$p(S^t|\Theta) = \sum_{\mathbf{z}} p(S^t, \mathbf{z}|\Theta) \tag{B.3}$$

Furthermore, assuming the outlier map $\mathbf{z}$ to be independent of the parameters Θ the complete data-likelihood $p(S^t, \mathbf{z}|\Theta)$ can be specified as:

$$p(S^t, \mathbf{z}|\Theta) = p(S^t|\mathbf{z}, \Theta)p(\mathbf{z}) \tag{B.4}$$

The first term in (3.23) can thus be reconstructed using the complete-process $p(\mathfrak{s}_j|z_j, \boldsymbol{\theta})$ defined in (B.1) and taking the negative logarithm resulting in:

$$\mathfrak{S}(\Theta, \mathbf{z}) = \sum_{j=1}^{N} -z_j \log p(\mathfrak{s}_j|\Theta) - (1 - z_j) \log \delta \tag{B.5}$$

This complete-process (inlier- and outlier-process) formulation is similar to the one defined in (A.6) leading to a joint estimation problem where besides the parameters Θ also the outlier map $\mathbf{z}$ is to be estimated. Following [76] and [73], the latent variable is out-integrated from (B.5):

$$\begin{aligned} \mathfrak{S}(\Theta) &= \textstyle\sum_{\mathbf{z}} \left[\sum_{j=1}^{N} -z_j \log p(\mathfrak{s}_j|\Theta) - (1 - z_j) \log \delta \right] \\ &= \textstyle\sum_{\mathbf{z}} \left[\sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j)) \right] \end{aligned} \tag{B.6}$$

where

$$h(x, \alpha(z), \beta(z)) = \alpha(z)x^2 + \beta(z) \tag{B.7}$$

with

$$\alpha(z) = \tfrac{1}{2\sigma^2} z \quad \beta(z) = -(1 - z) \log \delta + z \log \sqrt{2\pi}\sigma \tag{B.8}$$

such that:

$$\rho(x, \alpha, \beta) = \sum_{\mathbf{z}} h(x, \alpha(z), \beta(z)) \tag{B.9}$$

Out-integration comes down to computing the conditional expectation $E[x|y]$ of the complete data negative log-likelihood:

$$\mathfrak{S}(\Theta) = E[\mathfrak{S}(\Theta, \mathbf{z})|\mathbf{z}] \tag{B.10}$$

w.r.t the posterior $p(\mathbf{z}|S^t, \Theta)$. In other words, during integration the values of the latent variables z_j in the complete data negative log-likelihood (B.5) are replaced by the posterior $p(z_j|\mathfrak{s}_j, \Theta)$ values. From (B.4) we see that these posterior values conditioned on the observations $\mathfrak{s}_j$ are given by:

$$p(z_j|\mathfrak{s}_j, \Theta) = p(\mathfrak{s}_j|z_j, \Theta)p(z_j) \tag{B.11}$$

The solutions are obtained following a Mean Field strategy [108, 109] in which the posterior is approximated by the closest factorisable distribution

$p^f(z_j|\mathfrak{s}_j,\Theta)$. In our case this approximation $p^f(z_j|\mathfrak{s}_j,\Theta)$ can be set to a Bernoulli distribution $p^b(z_j|\mathfrak{s}_j,\Theta)$ over $\{0,1\}$ assigning a probability b_j to z_j being 1 and probability $(1-b_j)$ to z_j being 0. Minimizing the Kullback-Leibler (KL) divergence between $p(z_j|\mathfrak{s}_j,\Theta)$ and $p^b(z_j|\mathfrak{s}_j,\Theta)$ gives the Mean Field update equations. Note that these equations based on the prior outlier map distribution in (B.2) simply reduce to the Bayes estimate of b_j:

$$b_j = p(z_j = 1|\mathfrak{s}_j,\Theta) = \frac{p(\mathfrak{s}_j|z_j = 1,\Theta)p(z_j = 1)}{p(\mathfrak{s}_j|z_j = 1,\Theta)p(z_j = 1) + p(\mathfrak{s}_j|z_j = 0,\Theta)p(z_j = 0)} \tag{B.12}$$

Using (B.2) and (B.1) with $p^o(s_j) = \delta$, b_j becomes:

$$b_j = \frac{p(\mathfrak{s}_j|\Theta)P}{p(\mathfrak{s}_j|\Theta)P + \delta P_o} \tag{B.13}$$

such that the equivalent M-estimator in (B.9) becomes:

$$\rho(x,\alpha,\beta) = h(x,\alpha(b),\beta(b)) \tag{B.14}$$

Combining (B.14) and (B.8) it is straightforward to see that the parameters α and β of the equivalent M-estimator are dependent on the estimated σ of the inlier-process and fixed prior outlier knowledge parameters P, P_o and δ. In order to simplify the choice of prior outlier parameter values, we combine the three prior outlier parameters P, P_o and δ into a single prior outlier parameter λ according to:

$$\lambda = \frac{\delta P_o}{P} \tag{B.15}$$

such that

$$b_j = \frac{p(\mathfrak{s}_j|\Theta)}{p(\mathfrak{s}_j|\Theta) + \lambda} \tag{B.16}$$

and because $\log\lambda \approx \log\delta$ we replace the $\log\delta$ term in (B.5) with $\log\lambda$. Changing the parameter λ changes the amount of outliers and inliers. A higher λ suggests more outliers, while a lower λ suggests more inliers. A fine-tuning of this parameter is required in order to obtain the best result. In statistics, an observation $\mathfrak{s}_j$ is said to be abnormal with respect to a given normal distribution if the Mahalanobis distance $MD_j = \sqrt{\mathfrak{s}_j^2/\sigma^2}$ exceeds a predefined threshold. Equation (B.13) or (B.16) reflects the fraction b_j of a sample $\mathfrak{s}_j$ belonging to the inlier-distribution, called the inlier-belief. The outlier-belief can then be defined similarly but in terms of λ as:

$$1 - b_j = \frac{\lambda}{p(\mathfrak{s}_j|\Theta) + \lambda} \tag{B.17}$$

The outlier-belief exceeds the inlier-belief if $b_j < 0.5$ or $p(\mathfrak{s}_j|\Theta) < \lambda$, which is equivalent to $MD_j^2 > -2\log(\lambda\sqrt{2\pi}\sigma)$. Because of its dependence on σ, the Mahalanobis distance threshold above which $\mathfrak{s}_j$ is considered abnormal changes when σ is updated. Furthermore, observations are more easily rejected

from classes with a broad distribution than from classes with a narrow one, making the choice of λ dependent on the current registration problem and target surface, leading to very different λ settings for every different target surface to register. Because of these problems it is not clear how λ should be chosen. Ideally, a sample $\mathfrak{s}_j$ should be considered abnormal if $MD_j > \kappa$, where $\kappa \geq 0$ is an explicit Mahalanobis distance threshold that is equal for all normal inlier-distributions alike. Therefore, we replace λ according to [106] taking into account σ:

$$\lambda = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\kappa^2\right) \tag{B.18}$$

λ is now made inlier-distribution dependent and re-parameterized using the more easily interpretable prior outlier parameter κ. According to [106], the actual choice of κ can be regarded as the choice of a statistical significance level above which observations $\mathfrak{s}_j$ are considered outliers compared to the inlier-distribution. Figures B.1(a) and (b) show the equivalent M-estimator and outlier influence function respectively for the random latent variable modeling with $\sigma = 1$ and $\kappa = 2$. The resulting M-estimator resembles the Contaminated Gaussian (CG) estimator which is a continuous approximation of the Truncated Quadratic with better second derivative properties. Note that the width of the Gaussian within the CG in figure B.1(a) is approximately $\kappa\sigma = 2$ and thus dependent on the inlier-process σ and the prior outlier-process parameter κ. Furthermore, we see that the redescending of observation influences in figure B.1(b) starts approximately at $\kappa\sigma = 2$, which is of course related with the width of the gaussian. The observation influence becomes zero at approximately $\kappa\sigma = 4$. In figures B.1(c) and (d) we have plotted the equivalent M-estimators and influence functions for different values of $\kappa = \{1, 2, 3\}$. It is observed that the width of the gaussian in the CG and the point of redescending are altered accordingly. Note that the same M-estimators in figure B.1(c) could be obtained using $\kappa = 2$ with different $\sigma = \{0.5, 1, 1.5\}$ values, suggesting automatic adaptation of the M-estimator as a function of the inlier-process σ estimate during optimization. In other words, except for the choice of κ we have a self-regulating outlier rejection objective function. Due to the use of κ instead of λ as prior parameter, the outlier-process is coupled with the inlier-process, making the choice of κ easier compared to λ. This kind of random latent variable choice is used for the skull registration problem. Outliers within this registration problem are due to gross errors in the skull acquisition process and the construction and choice of the skull representation. These errors are treated probabilistically and a proper choice κ is to be determined. Further details are given in chapter 7.

B.3 Deterministic LV

Consider a binary latent variable z_j, but with very strong prior knowledge encoded in a deterministic (non-probabilistic) outlier detection function f that

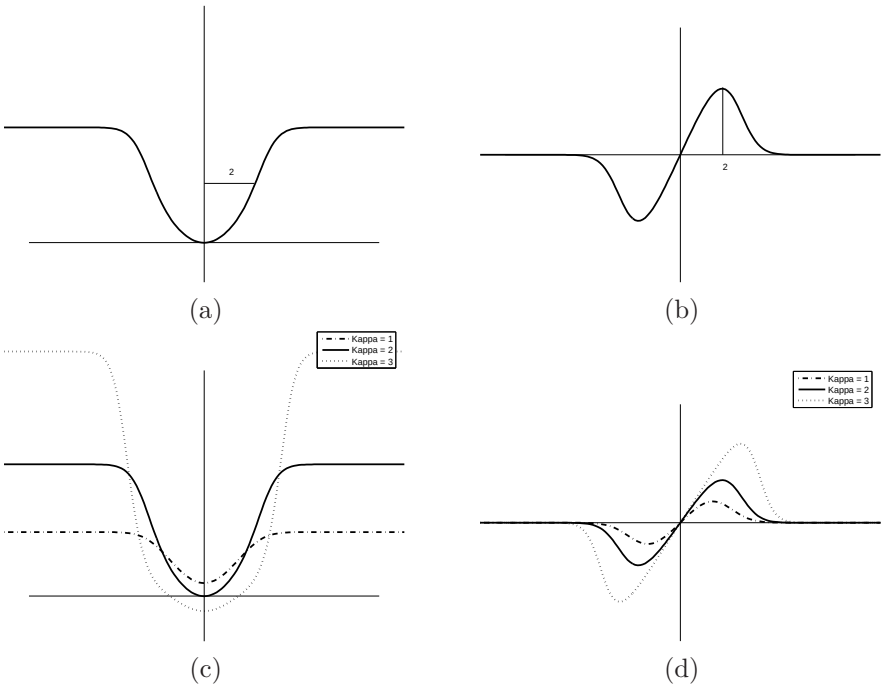


Figure B.1: Equivalent M-estimator and outlier influence functions for the random latent variable choice (a) and (b) $\sigma = 1$ and $\kappa = 2$. (c) and (d) the same for $\sigma = 1$ but with different $\kappa = 1, 2, 3$ values Note that the same M-estimators could be obtained using $\kappa = 2$ with different $\sigma = 0.5, 1, 1.5$ values , suggesting automatic adaption of the M-estimator in function of the inlier-process σ estimate

directly determines wether an observation $\mathfrak{s}_j$ is either an inlier or outlier:

$$z_j = f(\mathfrak{s}_j(\boldsymbol{\theta})) \in \{0, 1\} \tag{B.19}$$

The first term in (3.23) can again be reconstructed using the complete-process $p(\mathfrak{s}_j|z_j, \boldsymbol{\theta})$ defined in (B.1) and taking the negative logarithm resulting in (B.5). However, the latent variable estimation within the joint estimation problem simply reduces to a deterministic function evaluation instead of a minimization or out-integration. Starting from (B.5):

$$\begin{aligned} \mathfrak{S}(\Theta) &= \textstyle\sum_{j=1}^{N} \left[-z_j \log p(\mathfrak{s}_j|\Theta) - (1 - z_j) \log \delta\right]_{z_j = f(\mathfrak{s}_j(\theta))} \\ &= \textstyle\sum_{j=1}^{N} \left[h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j), \beta(z_j))\right]_{z_j = f(\mathfrak{s}_j(\theta))} \end{aligned} \tag{B.20}$$

where $h(x, \alpha(z), \beta(z))$ is the same as in (B.7) and (B.8), such that:

$$\rho(x, \alpha, \beta) = h(x, \alpha(f(x)), \beta(f(x))) \tag{B.21}$$

Analyzing (B.21) and (B.8) it is straightforward to see that the parameters α and β are determined based on the estimated σ of the inlier-process and the strong prior outlier knowledge encoded in the deterministic function $\mathbf{z} = f(\mathbf{x})$ besides the outlier PDF prior knowledge. Note that a resulting objective function using such a deterministic outlier detection function is completely self-regulating with regard to outlier rejection. This kind of deterministic latent variable choice is used for the patch registration and complete facial surface registration problems. Outliers within these registration problems are due to partial overlap between the target and floating surfaces. Individual point similarities defined in section 3.3.2 measured in floating surface points r_j^f not belonging to the partial overlap are very erroneous and therefore outliers. The deterministic function f generating binary valued latent variables can be defined based on the VBF $\mathfrak{b}(\mathbf{r})$ of the target surface. Further details are given in chapters 4 and 5.

B.4 Combined deterministic and random LV

A final choice of latent variable that we consider is not really different from the two previous, but rather a combination of them. Imagine an incomplete skull substrate is given for skull registration or patch, and facial surfaces with probabilistic outliers present in the partial overlap are given for the patch registration and facial surface registration respectively. To tackle these problems we define a combined latent variable outlier map $\mathbf{z} = \mathbf{z}' \wedge \mathbf{z}''$ with $\wedge$ a conjunction and $(1-\mathbf{z}) = (1-\mathbf{z}') \vee (1-\mathbf{z}'')$ with $\vee$ a disjunction incorporating a regular deterministic latent variable $\mathbf{z}'$ outlier map and a random latent variable $\mathbf{z}''$ outlier map according to:

$$\mathbf{z}' = f(\mathbf{x}) \tag{B.22}$$

and

$$p(\mathbf{z}'') = \prod_j^N P^{z_j''} P_o^{(1-z_j'')} \tag{B.23}$$

The first term in (3.23) (complete data negative log-likelihood) can then be reconstructed again using the complete-process $p(\mathfrak{s}_j|z_j, \boldsymbol{\theta})$ defined in (B.1) and taking the negative logarithm:

$$\mathfrak{S}(\Theta, \mathbf{z}', \mathbf{z}'') = \sum_{j=1}^{N} -[z_j' z_j''] \log p(\mathfrak{s}_j|\Theta) - [(1 - z_j') + z_j'(1 - z_j'')] \log \delta \tag{B.24}$$

The latent variable elimination in the joint estimation is now itself a two-step elimination. First the deterministic latent variable z' is eliminated:

$$\begin{aligned} \mathfrak{S}(\Theta, \mathbf{z}'') &= \textstyle\sum_{j=1}^{N} \left[-[z_j' z_j''] \log p(\mathfrak{s}_j|\Theta) - [(1 - z_j') + z_j'(1 - z_j'')] \log \delta \right]_{z_j' = f(\mathfrak{s}_j(\theta))} \\ &= \textstyle\sum_{j=1}^{N} \left[h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(z_j', z_j''), \beta(z_j', z_j'')) \right]_{z_j' = f(\mathfrak{s}_j(\theta))} \end{aligned} \tag{B.25}$$

where

$$h(x, \alpha(z', z''), \beta(z', z'')) = \alpha(z', z'')x^2 + \beta(z', z'') \tag{B.26}$$

with

$$\begin{aligned} \alpha(z', z'') &= \tfrac{1}{2\sigma^2}[z'z''] \\ \beta(z', z'') &= -[(1 - z') + z'(1 - z'')] \log \delta + [z'z''] \log \sqrt{2\pi}\sigma \end{aligned} \tag{B.27}$$

Then the remaining random latent variable z'' is eliminated:

$$\begin{aligned} \mathfrak{S}(\Theta) \quad &= \textstyle\sum_{\mathbf{z}''} \Big[\sum_{j=1}^{N} -[f(\mathfrak{s}_j(\theta))z_j''] \log p(\mathfrak{s}_j|\Theta) - \dots \\ &\quad [(1 - f(\mathfrak{s}_j(\theta))) + f(\mathfrak{s}_j(\theta)))(1 - z_j'')] \log \delta \Big] \\ &= \textstyle\sum_{\mathbf{z}''} \Big[\sum_{j=1}^{N} h(\mathfrak{s}_j(\boldsymbol{\theta}), \alpha(f(\mathfrak{s}_j(\boldsymbol{\theta})), z_j''), \beta(f(\mathfrak{s}_j(\boldsymbol{\theta})), z_j'')) \Big] \end{aligned} \tag{B.28}$$

Following the out-integration strategy earlier for the random latent variable we can define the equivalent M-estimator according to:

$$\rho(x, \alpha, \beta) = h(x, \alpha(f(x), b), \beta(f(x), b)) \tag{B.29}$$

As a conclusion for this section we will give the general complete MAP objective function from which we will start to solve the different registration problems given in the relevant chapters:

$$\mathfrak{E}(\Theta, \mathbf{z}', \mathbf{z}'') = \mathfrak{S}(\Theta, \mathbf{z}', \mathbf{z}'') + \nu \|\mathfrak{L}(\boldsymbol{\theta})\|^2 \tag{B.30}$$

with $\mathfrak{S}(\Theta, \mathbf{z}', \mathbf{z}'')$ defined in (B.24) such that after latent variable estimation the objective function becomes:

$$\mathfrak{E}(\Theta) = \mathfrak{S}(\Theta) + \nu \|\mathfrak{L}(\boldsymbol{\theta})\|^2 \tag{B.31}$$

with

$$\mathfrak{S}(\Theta) = \sum_{j=1}^{N} \rho(\mathfrak{s}_j(\theta), \alpha, \beta) \tag{B.32}$$

where

$$\rho(\mathfrak{s}_j(\theta), \alpha, \beta) = h(\mathfrak{s}_j(\theta), \alpha(f(\mathfrak{s}_j(\theta)), b_j), \beta(f(\mathfrak{s}_j(\theta)), b_j)) \tag{B.33}$$

Appendix C

Facial Thickness Acquisition

Adapted from

S. De Greef, P. Claes, W. Mollemans, M. Loubele, D. Vandermeulen, P. Suetens, G. Willems, Semi-automated ultrasound facial soft tissue depth registration : method and validation, *Journal of forensic sciences*, vol. 50(6), 7pages, 2005

C.1 Introduction

Besides facial surface information, tissue depths in 52 facial landmarks are acquired as well for every facial entry within the database. Relevant details, concerning thickness acquisition, are given in [1], which is a joint work of the authors. To make the thesis complete, I have chosen to give a summarized version of [1] in this appendix.

A mobile and fast, semi-automatic ultrasound (US) system is developed for facial soft tissue depth registration. The system consists of an A-Scan ultrasound device connected to a portable PC with interfacing and controlling software. 52 cephalometric landmarks are defined and the system is tested for repeatability and accuracy by evaluating intra-observer agreement and comparing ultrasound and CT-scan results on 12 subjects planned for craniofacial surgery, respectively.

We start with the description of the measuring system and an explanation of the validation. Then limited results are given and discussed after which we give a final conclusion concerning the Thickness acquisition. For further details and illustrations of the results we refer to [1].

C.2 Materials and Methods

Selection of Landmarks

Similar to the study performed by Brown et al. [110], the traditional landmarks of Kolmann and Büchly [111] were compared with previous studies on adult Caucasoid and other ultrasound studies. A total of 52 landmarks (LMs) were finally selected, 10 located on the midline and 21 located bilaterally. The selection of these landmarks was based on their presence in other studies (allowing comparisons) but also on the ability to reliably locate these landmarks in a standardized way on the face of the volunteers. An illustration of the landmark location is shown in figure C.1

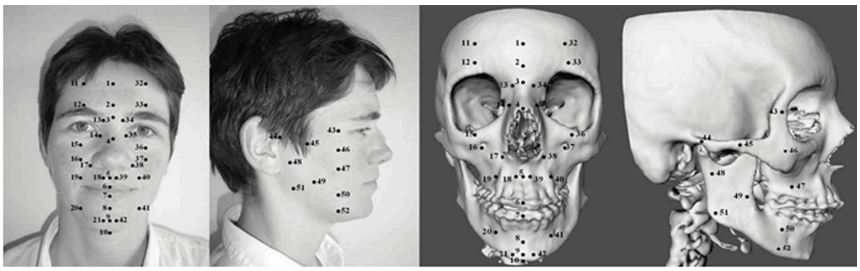


Figure C.1: Illustration of the landmark locations both on the skull and facial surface in a frontal and lateral view.

Measurement Device

In order to reach an as large as possible volunteer group and considering the pro's and con's of previous techniques, a mobile ultrasound (US) system was selected and software was developed to minimize overhead in data processing and storage. "A-scan"as well as "B-scan"ultrasound devices have been used in former studies. Because of the complexity of the "B-scan"image data, in contrast to the simple 1D curves of the "A-scan"-device, the increased storage requirements and data transfer times as well as the bulky transducers of "B-scan"-devices, an "A-scan"-device was selected (Epoch 4b^{TM} , Panametrics, Waltham, USA). This industrial ultrasound device is compact, mobile and lightweight (2.6 kg, battery included) . A flat and small (6 mm diameter), 10MHz ultrasound transducer is used for accurately pointing to and analyzing the landmarks. Furthermore, this device has a serial (RS-232) communication port, for connection with a PC.

Control Software

A Matlab^{TM}-based (The Mathworks Inc., Natick, MA, US) interface program was created to speed up the registration process by partially automating data

transfer and device setup. Except for the input of specific properties of the volunteers and positioning of the transducer on the landmarks, the whole process is controlled using a 3-button wireless infrared mouse. Mouse clicks control the following tasks: (1) freezing or unfreezing the A-scan curve; (2) transferring the curve from the ultrasound device to the PC, where the curve and the automatically calculated associated tissue depth are stored in the database; and (3) proceeding to the next landmark with automatic changing of the device settings. The tissue depths are calculated as the horizontal position (distance to US transducer) corresponding to the maximal peak within a predefined interval on the curve, called the gate. The settings of the Epoch 4b include the magnification of the measured ultrasound signal (the gain), the distance interval over which it is measured (the range) and the maximum peak detection window (the gate). These settings where originally based on the results of former soft tissue depth studies and refined after a training period with the "A-scan"-device.

Database Management

In order to easily extract specific data and perform statistical analysis, we created a database using MySQL software (MySQL AB, Uppsala, Sweden). The database is composed of 5 tables. One table contains all the personal information of the volunteer: name, birthday, health status, weight, length, facial profile, presence of dental prosthesis, plastic surgery, ethnic background, registration date and space for extra comments. A second table contains 36 different categories based on gender (M/F), age (18-35y, 35-50y, 50+), body mass index (-25, +25) and facial profile (concave, convex, straight). A third table contains the 52 landmarks of the face. Based on the volunteer's category and the landmark to measure, the interface program retrieves the necessary information, out of a fourth "settings"table, for the automatic adjustment of the Epoch4b^{TM} such as the range, gain and detection gate. Finally, in the fifth table, all the measurement results and corresponding curves are stored.

Registration Protocol

The registration procedure starts by entering the volunteer's identity and physical properties in the computer. The patient is measured in an upright relaxed position. With the probe as perpendicular as possible to the underlying bone and using a classic neutral coupling echo-gel, tissue depth is measured taking care not to indent the facial soft tissues. The correct transducer position is obtained by first compressing and depressing the tissues in order to differentiate the noise from the genuine US reflections. Secondly, the transducer orientation is interactively determined such that the highest peak, corresponding to the most perpendicular position of the transducer to the bone, is obtained. Finally, the compression on the soft tissue is maximally reduced avoiding losing skin contact. This technique was first practised on (fresh) corpses in combination with needle puncturing. Three measurements are obtained, the highest

Figure C.2: Thickness acquisition or registration setup

of which is taken into account for further statistics. As mentioned before, except for the correct positioning of the transducer, the data treatment occurs automatically using the mouse clicks (see figure C.2).

Validation: Repatability

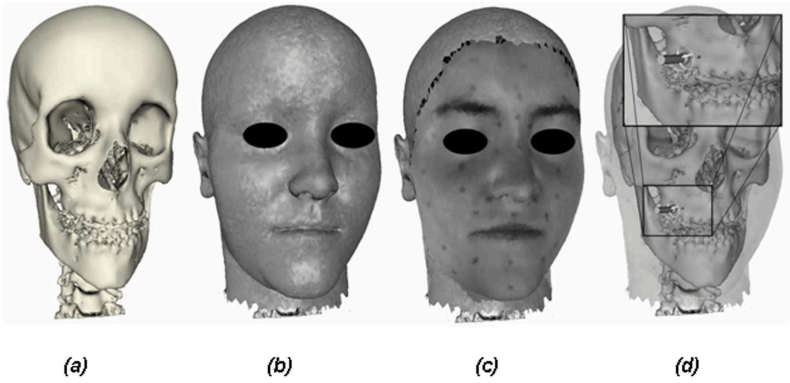


Figure C.3: Skull (a) and skin (b) surface segmented from CT scan image data. Textured skin surface (c) and virtual ultrasonography view (d)

For the repeatability evaluation of the US measurements a test group of 33 volunteers, composed of 19 males and 14 females with average age 39.0 years (s.d. 17 years) and average BMI 26.5 (s.d. 6.46), was measured twice (US1, US2), with time intervals varying between 2 days and 2 months. Accuracy was tested comparing ultrasound with CT-scan results. Twelve patients (11 females and 1 male with average age 19.7 years and average BMI 19.5) consented to have their facial soft tissue depths ultrasonically registered (US)

before acquisition of a total head CT-scan for preoperative osteotomy planning. Prior to the ultrasound registration, the 52 landmarks were marked on the face using a blue eyeliner pencil and a 3D picture of the face was taken using a 3D portable camera (ShapeCam, Eyetronics, Leuven, Belgium). The skull surface (figure C.3(a)) and external surface of the skin (figure C.3(b)) were extracted from the CT images by simple thresholding of the CT values (at a Hounsfield Unit value of 300 and -400, resp.). The extracted surfaces are represented at sub-voxel precision as a mesh of triangular tiles using standard surface meshing software. The 3D facial surface obtained with the 3D camera was automatically fitted to the CT-based skin surface (minimizing the distance between the two surfaces using the method). This allows texturing of the CT extracted skin surface with the texture obtained from the 3D picture (figure C.3(c)) in order to determine the CT-coordinates of the exact landmark locations as measured with ultrasound. A software program was developed to perform virtual "A-scan"ultrasonography (figure C.3(d)), allowing to measure the CT-based soft tissue depths at the 52 different landmarks locations as indicated by the blue points prior to ultrasound registration. The virtual US probe is manually positioned on a 3D graphical rendering of the textured CT skin surface at each skin-based landmark location. The algorithm then iteratively estimates the corresponding skull base point at which the normal to the skull surface is as parallel as possible to the normal to the skin surface at the skin landmark. This mimics the real ultrasound measuring protocol as closely as possible. The distance between the skin landmark and corresponding skull base point is stored as the associated CT soft tissue depth. Finally, in order to examine the influence of gravity on the differences in tissue depths as measured by CT and ultrasound, the external skin surface extracted from CT, measured in supine position, was aligned for each individual with the 3D facial surface obtained by the 3D camera in upright position. In order to limit the influence of expected differences, the surface alignment is only based on corresponding surface parts that are less than a maximal (2mm) distance apart.

Statistical Analysis

The intra-observer agreement was statistically analyzed with a paired t-test and a Wilcoxon paired Signed Rank test. The latter procedure tests for the null-hypothesis of the median of the paired differences to be zero. The accuracy of US measurements compared to CT-based measurements was statistically analyzed using the Wilcoxon paired Signed Rank test. Given the limited ($n = 12$) number of observations, this non-parametric test was chosen since it has a higher power efficiency (higher sensitivity for smaller sample sizes) and higher robustness to violations of the normality assumptions. Confidence intervals for the median of the differences were calculated for each landmark using a bootstrapping technique (resampling with replacement). Bland-Altman graphs are constructed for selected landmarks to picture the relationship between differences, average values and confidence limits. A robust linear regression between CT and ultrasound measurements was calculated to look for

any linear bias (offset and/or slope) in either the ultrasound (time-to-distance transformation based on estimated sound velocity in soft tissue) or the CT segmentation thresholds. All tests were implemented using the Matlab^{TM} 6R13 (The Mathworks Inc., Natick, MA, US) data analysis software.

C.3 Results and Discussion

Figure C.4: Absolute distances between 3D skin surface prior to US acquisition (upright position) and 3D skin surface extracted from CT image data (in supine position). The distances are shown here for one particular subject.

The repeatability study shows very few ($n = 3$) landmarks with a statistically significant ($p < 0.01$) difference between the repeated runs. A closer look at the protocol application during this first stage of the project indicated that a slight change of transducer position had occurred between the first and second measurements. Indeed, the position of the supra-orbitalis landmark changed from 'on the eyebrow' to 'just above it' because of the eyebrow interference. This probably explains the significant differences for LMs 12 and 33. Furthermore, the median difference for LM 32 is very small ($< 0.5mm$) compared to the corresponding average thickness (4.6 mm). The slope coefficient ($a = 1.01$) of the linear regression between CT and US measures, is very close to 1. Any miscalibration of the ultrasound device, by an improper setting of the sound velocity (set at 1452 m/s) in the measured soft tissue, would result in a value substantially different from one. We can thus conclude that this setting is accurate enough, relative to the CT standard used. The intercept value ($b = 0.23$) is close to zero. Any improper setting of the thresholds used for defining the external skin and skull surfaces in the CT images, would result in a constant offset relative to the US measures. Since the CT images were acquired and processed with typical voxel dimensions of 0.5 mm x 0.5 mm in-plane and 0.9 mm transaxially, we can conclude that the thresholds chosen are sub-voxel precise, on average. Careful examination of

the Bland-Altman graphs for all landmarks shows a small subset of clear outliers both for CT and US based measures. However, upon visual inspection of these graphs, no clear preference in any direction (positive or negative) or type (CT or US) could be observed. Note that these outlier values do not influence the results drawn by the Wilcoxon paired signed rank test, which is known to be robust to outliers. Their existence, however, requires robust statistics to be used when further processing the database. One practical consequence of this finding is that, during acquisition of the US data, a left-right check on the values obtained is calculated and the operator is given the opportunity to reregister certain landmarks signalled to be outside an acceptable left-right difference range ($< 3mm$). Since landmarks found to be statistically different between CT and US were all located in the masseter region, we examined whether the differences could be explained by the influence of gravity on the soft tissue thicknesses between the upright position during the US acquisition and the supine position in the CT-scan. Figure C.4 shows the typical pattern as observed for an individual case. A zero-difference (dark) band running over the naso-labial sulcus can be observed. When inspecting the aligned CT and 3D skin surface, the thickness of the regions in front of this zero-band seems to decrease from 3D to CT whereas the opposite effect occurs for the region behind. Although a relatively weak ($r = 0.22$) linear correlation could be found between the (CT versus US) errors in the landmarks and the (3D minus CT) distances between the skin surfaces in the upright and supine position, paired per landmark and per subject, a more pronounced linear correlation ($r = 0.53$) could be established between the median differences and distances. We see that the larger (CT minus US) differences and (3D minus CT) distances are located precisely at the landmarks in the gonion-, supraglenoid-, and occlusal line region, substantiating the hypothesis that they are probably due to the gravity effect.

C.4 Conclusion

Thanks to the progress in computer science and medical imaging technology it was possible to create a fast, mobile and user-friendly facial soft tissue depth acquisition system. Statistical analysis of the repeatability and accuracy proved our system to be a reliable and accurate measurement tool. A correct application of the protocol allows in 20 minutes the measurement of 52 facial landmarks in a non-invasive, standardized and repeatable way. It will allow (re)evaluation of older facial soft tissue depth data based on a large group of subjects of different age, sex, race and build. An evaluation that shouldn't be limited to the facial soft tissue depth data but, as the actual tendency, expanded to other facial features in order to increase the degree of accuracy of the facial reconstruction.

Appendix D

Patch registration and integration: An overview

D.1 Introduction

Not to overload chapter 4 and to keep the focus on the main application of the thesis: craniofacial modeling and reconstruction, we have chosen not to incorporate an exhaustive related work review concerning patch registration and integration. The most relevant related work is referred to within chapter 4, while in this appendix a more elaborated review is given. For clarity, a model in patch registration and integration literature is the final result or a complete 3D acquisition of an object's surface.

Mentioned in chapter 4, systems for measuring three dimensional data have in common that for building complete models, some degree of human intervention or a priori knowledge about the geometry of the imaging process is necessary. The lack of fully automatic systems without any kind of a priori knowledge or user intervention makes it very labor intensive and expensive to make databases of 3D models. Most Three-dimensional model acquisition systems generate several partial reconstructions that have to be combined for building complete 3D models. The reason is that the acquisition system can only compute 3D information of what can be "seen"from a certain viewpoint. We refer to the reconstruction from a single viewpoint as a patch. A complete model is build from many patches and can be seen as a patchwork. The number of views necessary is mainly determined by the complexity of the object. With a more complex object, more views are needed and it becomes interesting to have a fully automatic reconstruction system.

Combining several patches into a single model traditionally, involves two main phases. First the patches need to be aligned into a common coordinate frame, the registration phase. Secondly, the registered patches need to be integrated into a single entity, which is the integration phase. The registration phase itself can be considered as two different problems that need to be solved

[81]. The first problem is a global registration problem where no a priori information on relative positions and orientations of the patches is available and the second is a pose refinement, where it is assumed that an initial estimate of the exact positions is known. We refer to the former problem as *crude registration* and the latter as *fine registration*. When more than two patches are to be registered, both crude and fine *pair-wise* registrations need to be combined into a *multi-view* registration.

D.2 Fine registration

The most famous fine registration algorithm is the iterated closest point algorithm (ICP) introduced in [59], where the distance between corresponding points is iteratively minimized. More or less at the same time, a very similar algorithm was introduced in [112]. The difference is that [112] minimizes the distance between a point and the tangent plane in its closest point partner. The method in [112] is often referred to as point-to-plane ICP and [59] as point-to-point ICP.

The basic ICP algorithm has several problems and has been improved by many authors. Firstly, when aligning point sets, they are typically two different samplings of the same shape. Even when the patches are perfectly aligned, the distance between a pair of best matching points can be nonzero. A possible solution is to look for corresponding points in between the available points, by triangle interpolation for example when a mesh representation is available. This problem was solved in chapter 4 using continuous implicit target patch representations.

Secondly, the original algorithm in [59] was proposed for registration of point sets with complete overlap. When dealing with partial overlap, many points of one patch will not have a corresponding partner on the other patch. However, points without partner do get matched to some point on the other patch. Such incorrect matches are referred to as *outliers*. Dealing with outliers can be done using two different strategies. The first approach is to replace the function that is minimized , e.g. [113] using a LMSQ error function instead of a classical LSQ estimator and [114] using M-estimation. The latter is similar to our approach. The second approach is to identify the outliers and to remove them from the list of matches, which is similar to our deterministic outlier detection function. Both strategies are combined within our outlier handling of section 3.3.3, through the use of a combined latent variable incorporating a deterministic regular latent variable and a probabilistic random latent variable. A study of outlier classification (second strategy) is presented in [115]. The simplest way to detect outliers is using a distance threshold [116, 117, 118]. In [119] a new outlier distance threshold is calculated in every iteration step, by analyzing the statistics of the distances between the point pairs. A similar approach is done in [120]. In [121] a heuristic approach is used and the distance threshold is set to $t = cs$, with s the range scanner resolution and c an empirically determined value based on the distance between

the center mass of the points sets. Many other criteria have been used for rejecting outliers [116, 117, 122, 123, 124, 125] and multiple criteria can be combined [126]. Still, most outlier detection systems rely on the distance from the floating patch to the target patch, which is dependent on the relative pose between the two patches. Our deterministic outlier detection function does not suffer from that shortcoming, because it is solely dependent on the target patch.

The third problem of the basic ICP is the very expensive closest point, being corresponding points, search. A naive closest point search requires $N^f * N^t$ point-to-point distance computations during every iteration, where N^f are the number of points on the floating patch and N^t are the number of points on the target patch. Several methods have been developed to speed up the closest point search. In [117, 127, 128, 129] K-D trees are used to speed up the search. A simple but more popular method is to sub-sample the floating patch points. This sub-sampling can be randomized [126, 130] or geometric. In [120, 131] the principal of inverse calibration is used to speed up the search and in [132] ICP is accelerated using a multi-z-buffer technique. This problem was solved in chapter 4 using continuous implicit target patch representations, without the need for explicit point correspondences during optimization.

An alternative fine registration method is given in [133], which is based on the maximalization of mutual information and an excellent survey, about fine registration, can be found in [134].

D.3 Crude registration

The most important problem with fine registration methods is that they perform an iterative optimization and as such are sensitive to local extremes. For convergence, the algorithm needs a rather good estimate to start from. Due to convergence dependency on the initial pose, fine registration is sometimes referred to as *constrained registration*. Finding relative poses with arbitrary differences in initial pose is referred to as *unconstrained registration* or *crude registration*. Typically a coarse-to-fine strategy is used [80, 135]. First, a crude registration is performed to give an initial estimate of the transformation. Next, a fine registration is performed, computing an accurate transformation. Crude registration is successful as long as it gives an accurate initialization for the fine registration. Having a more accurate crude registration does have the advantage that the fine registration can converge more robustly and within fewer iteration steps.

Crude registration is often solved manually or with a priori knowledge about the geometry of the imaging process. The former is done by interactive indication of point correspondences [120] or manual rotation and translation of the patches. This is a common approach in many commercial 3D acquisition systems. The latter is done mechanically [118, 135, 136, 137, 138] or based on calibration information of the scanner [139]. However, these are scanner dependent and sometimes additional mechanical tracking hardware is required.

Two different approaches for automatic crude registration without knowledge about the 3D sensor are possible, a pose search approach and a correspondence search approach. [59] already suggested an exhaustive *pose search* method to initialize their ICP fine registration algorithm. In [140] the same idea is implemented in parallel so that different possibilities can be checked at the same time. In [141] a method is used based on the RANSAC principle. They introduce a data-aligned rigidity-constrained exhaustive search (DARCES) algorithm. However, such exhaustive approaches require huge computational efforts. Less exhaustive pose search algorithms make use of global descriptions of the surfaces. The idea is to compute global properties that allows the recovery of the initial relative pose between patches. For example, the center of gravity can be used to estimate an initial translation between patches. In [142] a frequency domain pose search is performed, where texture information is incorporated in order to disambiguate among different possibilities Another global patch representation is the extended Gaussian Image [143, 144]. However, when the amount of overlap between patches decreases, the estimated transformation based on global descriptors and properties is less robust and/or accurate enough.

For crude registration of partial overlapping patches it is advisable to use a *correspondence search* approach based on local descriptors. A first category of local descriptions, use special points, curves or regions and describe these special structures with several parameters or invariant descriptions. Corresponding structures in between different patches are searched for in order to determine the initial transformation. In [145] curves of extreme principal curvature are calculated. In [146], bitangent point pairs are used and [80] extended them to bitangent curve pairs. In [147] curves of constant mean and Gaussian curvature are used. A problem with these special structures is that enough overlap in between different patches is needed in order to guarantee the presence of these structures. Sometimes objects don't have the special structures at all, limiting their use.

An alternative category to determine corresponding points is to calculate *point-signatures* or *point statistics* for a large amount of points on every patch. In general, point-signatures are descriptions of the local geometry around a point on the surface. By comparing point-signatures of different patches, possible correspondences can be established, because corresponding points will generate similar local geometry descriptions. Through comparing point-signatures the 3D matching problem is often reduced to a 2D or 1D matching problem. In [148] the intersection of a patch with a small sphere centered in a patch point is computed. The intersection is a 3D curve, which is projected onto the tangent plane of the point. The distance between the 3D curve and its projection is used to create a 1D signature. In [149] a set of 2D contours is used, which are the projections of geodesic circles onto the tangent plane, where the normal variation along the geodesic circles is also encoded. [135, 150, 151] propose spin-image point-signatures, which are in fact local 2D geometry histograms. One parameter of the histogram is the distance to the tangent plane of the center point from all the surface points. The other

parameter is the distance to the normal of the center point. [135, 151] uses spin-images for object recognition and [150] applies them for surface registration. In [152, 153] a modified version of the spin-image is proposed, which is called the surface point signature. One parameter of the histogram is the distance between the center point and every surface point. The other parameter is the angle between the normal of the center point and the segment created by the center point and every surface point. Instead of using point-signatures one can also define triangle-signatures [154]. For all triangles on a patch a 2D histogram describing the relative position of the neighboring triangles is computed. In [155] a token set or feature vector is attached to each triangle.

Once point-signatures are calculated for points on every patch, a set of possible corresponding points is established through point-signature comparisons. However, it is impossible that every corresponding pair is correct, because of the partial overlap between patches. It is thus important to determine the transformation out of the correspondences in a robust manner, which is able to distinguish correct from incorrect correspondences. A simple and naive way is to select the best correspondences based on the comparison value of the point-signatures. In [89] groups of correspondences are formed, according to geometrically consistent conditions and the distance between correspondences. This method is already better than the naive one, but it still fails when having a small amount of overlap, because the grouping criterium rejects different correspondences that are to close to each other. Having a limited amount of overlap all good correspondences are grouped close to each other, so the criterium has to be fine-tuned towards the amount of overlap. A robust and elegant strategy is to use directed point-signature in combination with a voting scheme [83]. A directed point-signature is a point-signature combined with a directed component (e.g. the surface normal). All possible rigid transformations that translate a point to its corresponding point and that makes their normal directions coincide is calculated and gets a vote in a voting scheme. This is done for every corresponding pair and the transformation that gets the most votes is selected.

D.4 Pair-wise versus Multi-view

So far only fine and crude registration algorithms between a pair of patches were discussed. Finding the relative pose of two patches is referred to as *pair-wise registration*. When more than two patches have to be registered, a *multi-view registration* is needed. In this case the absolute poses of every patch into a common coordinate frame, instead of the relative poses, is searched for. Again a distinction between multi-view crude - and multi-view fine registration is made. During multi-view registration some new complications are introduced. A multi-view registration algorithm is often based on pair-wise registrations, where the common strategy is to do consecutive pair-wise registration steps. However, (1) which patch pairs will be registered? Patches with no overlap should not be registered. Further complications are (2) the building up of

registration errors and (3) inconsistent poses. In a crude registration context, the former two complications are relevant. In a fine registration context, the latter two complications are relevant.

Multi-view crude registration in [117] is solved by tracking the scanner in combination with an interactive alignment. [82] and [81] also use scanner knowledge to start from. In [156] the information about the sequence in which the patches were recorded is used in order to know which pairs of patches are to be registered. These methods as such are not considered fully automatic without prior knowledge. In [157] a trial and error approach is performed in some way corresponding to the way a human solves a jigsaw puzzle. All possible patches are registered with a pair-wise registration based on spin-images. A global solution is then found by evaluating the quality of pair-wise registrations. Measures are introduced that allow local and global consistency checks. All patches are registered such that a global consistent solution is found based on local consistencies. In [80] a similar approach is taken, where the quality of pair-wise registration is expressed in terms of the number of transformation inliers. Subsequently, a minimum spanning tree is constructed to select a minimal set of transformations that transform every patch into a common coordinate system.

In a *multi-view fine registration* context the error build up and the inconsistent poses are the most important complications when using pair-wise fine registrations and have been analyzed by several authors [112, 117, 131, 132, 158, 159, 160, 161]. Accuracy improvement has been studied by minimizing a global error function. Another strategy besides a combination of consecutive pair-wise registrations is perform a simultaneous registration of a floating patch with every other patch being multiple simultaneous target patches . In [128] the closest point on all the target patches is looked for. However, an severe closest point computation cost increase is required. Other examples of simultaneous registration are given in [158, 162, 163, 164] not suffering from error-build up, but the complexity grows exponentially with the number of patches. Instead of registering a floating patch with every target patch simultaneously one can also iteratively register every patch with an intermediate integration of all the target patches until convergence [81, 82]. A final strategy is to register patches sequentially to previously registered and integrated patches as in [81], which is a good compromise between consecutive pair-wise registrations and simultaneous registrations.

D.5 Integration

Surface *integration* methods differ by the type of input data used, unorganized or connected point sets, and the type of surface representation, parametric or implicit. Examples of methods integrating unorganized point sets are [165, 166], using parametric surfaces, and [167, 168], using implicit surface representations. Since these methods do not require a specially organized input set, they can be applied in more general situations. At the same time,

however, these methods are less robust against noisy data, outliers and cannot reliably integrate high curvature regions. The integration can be improved using structured input data and parametric surfaces [169, 170], but according to [85], they can still fail in areas of high curvature. The more successful approaches use structured data, while hiding the topological problems in the previous parametric surface based methods by using implicit volumetric representations. Johnson et al. [89, 90] create surface occupancy grids, which were the earliest and simplest volumetric representations. However, final surface extraction, based on ridge-detection in the surface likelihood, is not robust [85]. In [85, 171, 172], volumetric integration algorithms are presented that construct a grid-based weighted signed distance function to the final surface. Triangular surface representations or mesh tessellations are extracted using Marching Cubes (MC) [87]. Interesting properties of volumetric integration using signed distance functions on a grid, (e.g outlier robustness, incremental update, incorporation of range uncertainty, ...) are listed in [85].

The algorithms using signed distance functions on a grid differ in the way the implicit surface is constructed and the volumetric data is organized. A disadvantage of the ray shooting algorithm used in [85] for signed distance grid building, is that the signed distance functions are dependent on the sensor location. Other authors [82, 171, 172] resolve this problem by constructing the distance functions on the grid differently. It is also preferable to incorporate patch normals during build up of the distance functions in order to cope with sharp structures in the model. A major drawback of these volumetric integration methods is the use of a 3D grid, such that the final model resolution is limited by memory requirements and computational complexity. In [85] this problem is alleviated by run length encoding of each 2D slice in the grid, which depends on a complicated space carving procedure. In [172] grid octree structures are used, which is a hierarchical representation capable of increasing grid resolution locally. In [91] the octree idea with hierarchical space carving is combined. Whether an octree or space carving is used, both methods introduce additional overhead, like grid-point classification and octree building, in order to cope with the memory problem. Furthermore, features smaller than the grid sampling size will be missed.

Signed distance functions are superior to occupancy grids because integration and multi-view fine registration can be unified into a single step, while traditionally they were performed separately. In [81] and [82] an implicit signed distance field is used in order to have a common representation for both integration and registration, where the gradient of the distance field is used to determine closest points for registration. The advantage of the volumetric representation construction used in [81] compared to [82] is the use of filtered normals, such that their algorithm is more robust against noise in the reconstruction data. A disadvantage of both [81] and [82] is that a memory inefficient and expensive grid is used as a discrete sampling of the signed-distance field, which limits the final resolution and registration accuracy. This is a drawback to all volumetric registration and especially integration approaches till now.

It still remains unanswered to what extent signed distance build-up is sen-

sitive to the presence of noisy data. Is the final signed distance value in each grid-point reliable, when the input patch is very noisy (larger than the grid sampling size,e.g.) and does this affect the final integration? Especially patch normal information is severely influenced by noise. Building up the signed distance function with incorporation of these corrupted normals will lead to a wrong result.

Every surface integration method also has to cope with the presence of holes in some of the patches. A possibility is to perform hole filling after the integration and final mesh extraction, which is difficult when the holes are highly non-planar. In [85] hole filling is performed on the volume representation, but special grid-point classification into near-surface, empty and unseen is necessary. In [173] holes in the distance field are filled by fitting local quadrics. No volumetric integration method so far uses a hole filling inherent representation.

The amount of literature concerning texture integration or blending is limited compared to shape integration. In [89, 90] texture blending is accomplished using weighted averaging of overlapping textures from the original contributing patches. Texture weights are dependent on the angle between the consensus surface normal and the viewing direction or relative pose of the patch. Pulli *et al.* [91] also perform a weighted texture blending, where the weights depend on the angle between the viewing direction of the contributing patch and the viewing direction of a virtual viewer, such that texture integration is view-dependent. Our texture integration is very similar to the shape integration using VIF to represent color information. Doing so a viewing independent texture integration is obtained. Disadvantages of current texture integrations methods [89, 90, 91] are the need for surface extraction before texture weight determination and the dependency of the weights on the relative poses of the patches, which makes the texture weights and blending sensitive to the accuracy of the patch registrations. Furthermore in [89, 90] additional vector information is stored for every point of the grid, thereby further increasing the memory requirements.

We constructed a volumetric shape and texture integration technique in section 4.3. The shape integration is intertwined with the multi-view fine-registration in section 4.2.4. To cope with problems of current volumetric integration techniques enlisted in the previous paragraphs within this section we used variational implicit functions and VIS representations instead of implicit grid representations. Doing so, several improvements were gained, e.g. hole filling, dealing with noise, no expensive grid, no sensor dependency, ...

D.6 Conclusion

As a conclusion, four different registration tasks (Crude/Fine pair-wise registration and Crude/Fine multi-view registration) and an integration task are to be solved for complete surface acquisitions from partial surfaces. In order to have a fully automatic registration and integration, without any kind of a

priori knowledge or user intervention, more than one registration task needs to be solved. First a global registration needs to be solved using a crude registration. Then a pose refinement is performed with a fine registration. When more than two patches are to be registered, both crude and fine pair-wise registrations are to be converted into a multi-view registration, where additional complications are present. Finally the registered patches are to be integrated into one entity. One can choose the best algorithm available for every task and combine them to have a fully automatic method, but it would be interesting to have a common representation of the patches used, while solving every task, like the combined multi-view fine registration and integration in [81, 82].

Our complete automatic registration algorithm in chapter 4 for aligning multiple partially overlapping surfaces for the surface patches consists of a coarse-to-fine strategy as in [80] resulting into four stages. First, a robust but less accurate pair-wise crude registration between every pair of patches is calculated based on point correspondences, which are established by comparing directed point-signatures calculated from the VIS representations. A pose clustering procedure [83] selects the best rigid transformation from valid transformations that map corresponding point pairs, while making their directed (normal) components coincide. Local 2-D geometry histograms calculated from the VIS representations are used as point-signatures, which are an improved variant of the spin-images in [84]. Second, a selection of all pair-wise crude registrations are combined into a multi-view crude registration by using a minimum spanning tree algorithm (MST) like [80]. We define the pair-wise transformation cost in terms of the number of inliers supported by the transformation. The result of the MST algorithm is a non-redundant set of relative pair-wise transformations that transform all patches in the same coordinate system. This set of initial transformations is further improved, in a third step, by applying pair-wise fine registrations, which are less robust in terms of initialization but more accurate. These fine registrations are calculated iteratively by a gradient descent minimization of a distance error criterion, based on the continuous VIS representation. In order to deal with partial overlap between patches a novel deterministic outlier detection function is defined, which is independent of the relative pose or distance histograms between two partial surfaces. Both the minimization and outlier handling are defined and combined using the theoretical robust surface registration framework from chapter 3. Finally, in a fourth step, we cope with pair-wise registration error accumulation by applying a simultaneous multi-view fine registration of all partial surfaces with intermediate integrations. The fine registration of a patch with an intermediate integrated surface is similar to the pair-wise fine registration. For the integration we use an implicit volumetric method. This integration also uses the VIS representation, such that a common representation is used throughout the whole reconstruction pipeline.

Appendix E

Non-rigid surface registration: An overview

E.1 Introduction

In order to perform a statistical shape analysis or to construct a shape model, dens inter-subject correspondences in between facial surfaces are needed. Therefore, the 3D points with known connectivity of a carefully constructed generic reference face are mapped or registered non-rigidly onto the faces in the database. The generic reference face is the floating surface while the faces in the database are the target surfaces to which the floating surface is registered. A non-rigid transformation during registration is required due to the local differences in geometry of both surfaces. The establishment of dens point correspondences using a non-rigid surface registration in between two surfaces is fairly complicated due to the the non-rigid nature of the registration problem. Furthermore, it is a well known fact that the quality of the statistical model extracted afterwards is dependent on the quality and the consistency of the established dens point correspondences.

The facial surface registration methodology must also be robust against noise and must be able to cope with outliers due to partial overlap or missing data. The latter is because of the fact that the 3D surface acquisition device used in chapter 4 is not able to acquire 3D information from hairy or covered regions in the face, generating incomplete 3D surface acquisitions with holes or missing data. The former, or noise robustness, is not necessarily due to small errors in the target surface but rather due to the non-rigid nature of the registration. An instance of point correspondences during registration is typically erroneous, especially at the beginning of the registration, because the shapes of the floating and the target surface are then very different. In the following section a very short overview is given of non-rigid 3D surface registration algorithms for shape-modeling. Note that alternative non-rigid registration algorithms exist (e.g. [174, 175]), but these are not designed to

deal with surfaces and are therefore not incorporated within the overview.

E.2 Short overview

In [176], the dependency of the shape model on the correspondence quality is explicitly exploited. Dens correspondences are searched for, while minimizing the statistical model description length, being a model quality measure. Floating and target surfaces are represented using parametric representations and dens correspondences are defined using re-parameterizations of the surfaces. However, the optimization is not straightforward and a complicated genetic optimization algorithm is therefore applied. Furthermore, severe surface preprocessing is necessary as the technique can not cope with local surface inconsistencies and missing data.

[177] describes the surfaces using features based on curvature information, but then the shape of the floating an target surface may not differ to much. The same problem is encountered in [178, 179], where the floating surface is first rigidly aligned with the target surface, after which a brushfire algorithm is initialized to fill in dens correspondences.

Several authors cope with this floating and target surface shape difference problem by simplifying the non-rigid registration. The technique in [94] is similar to [178, 179], but user-intervention is incorporated to guide the registration. Both [95] and [96] place manually a limited amount of corresponding landmarks (9 and 15, respectively) in between the floating and target surface, based on which an initial TPS non-rigid transformation is determined and applied in order to bring the geometry of the floating surface closer to the geometry of the target surface. Afterwards this new situation is used as a starting point to derive dens correspondences being closest points in [95] and using a coating procedure in [96]. However, the final result is then dependent on the quality of the manual landmark indication.

An alternative simplification is to redescribe the 3D surfaces into 2D height maps or images , based on a cylindrical 3D-2D mapping [97, 98]. The registration problem is then solved in 2D using optical flow [97] or a RBF transformation with additional user indicated landmarks [98]. However, a 3D-2D mapping incorporates the danger of loosing 3D surface information, especially in regions with high curvature information. Therefore, we prefer to solve the problem in 3D instead of a lower dimension.

A fairly complicated but interesting geodesic surface matching framework is proposed in [180] and was first adopted by us and used in [7] for example. During geodesic surface matching, corresponding points are points connected through a minimal cost path (geodesic path), going from the floating surface to the target surface, on a 4D cost surface, where the cost is determined based on Euclidian distance and local curvature similarity between the end points of the path. However, we experienced that a proper shape initialization was required, such that we opted for a manual landmark placement generating an initial TPS transformation in [7] like [95] , making the result dependent on

the landmark qualities. Furthermore, the 4D cost function is defined on a volumetric grid, such that the final resolutions and accuracies are dependent on memory requirements.

The non-rigid robust point matching (RPM) algorithm [71], was used to find initial landmark correspondences in between the 52 indicated face landmarks of the faces, in which ultrasound thicknesses were measured, to determine an TPS transformation initializing the geodesic surface matching in [7]. The RPM algorithm basically applies two techniques to tackle the non-rigid registration problem in between two point-sets, soft-assign (for correspondence establishment) and deterministic annealing (for a TPS transformation update). The reason for the RPM algorithm being able to cope with arbitrary initial geometry difference is due to the deterministic annealing optimization. In theory the RPM algorithm could be directly applied for the surface matching problem by representing both floating and target surface as dens point-sets. However the original RPM cannot handle a huge amount of points, due to the size of the soft-assign matrix which is dependent on the amount of points and due to a direct implementation of the TPS deformation. Furthermore, a point based target surface representation leads to sampling errors within the final correspondences similar to the ICP algorithm for rigid patch pose refinement. The RPM is closely related with the non-rigid surface registration algorithm in chapter 5. The same deterministic annealing optimization is applied to find a TPS based non-rigid transformation. However the TPS transformation calculations and evaluations are feasible for a huge amount of points, by making use of the VIF machinery in chapter 3. Furthermore, the target surface is represented using a VIS, such that no sampling errors are contained within the final correspondences and such that the establishment of corresponding points during registration for a huge amount of points is still computationally feasible.

Appendix F

Continuous numeric gender values

F.1 introduction

Principal component analysis requires numeric values for every facial information part in the vectorial representations. In contrast to BMI and age it is not straight forward to obtain a numeric value for the gender property. A possible solution consists of assigning a discrete numeric value to males (e.g. + 1) and females (e.g -1). However, some facial entries have more pronounced male/female facial features compared to other facial samples in the database. It is well known fact that the difference between males and females is more pronounced during aging. Younger persons constantly evolve starting from a genderless childish outlook. Middle aged people are completely evolved showing apparent gender differences. So instead of a discrete numeric value a more continuous value is preferred in order to differentiate between the male and female alikeness of faces. The more positive/negative a gender value the more male/female a face appears to be. A value of zero corresponds to a genderless looking face.

A potential way to obtain a continuous gender value is to define a numeric scale and to manually score every facial entry in the database accordingly. However, this is quite labor intensive and very subjective. Instead we propose an automatic and objective continuous numeric gender value estimation using a classifier based on the support vector machine methodology. Classification learns a decision function associated to a set of labeled data points to predict the values of unseen data. Support Vector Machines (SVM) is a powerful methodology for solving problems in classification [181], function and density estimation.

F.2 SVM Classification to obtain continuous values

The classification in our case is a binary (male/female) classification. This classification divides two separate classes, which are generated from training examples. Each example consists of data points combined with a binary label +1 or -1. Using the SVM methodology and based on the training data the two classes are separated by introducing and computing a separating hyperplane, which must maximize the margin between the two classes. This the optimum hyperplane minimizing the distance between the closest +1 and -1 data points, which are known as support vectors. The decision whether an unseen data sample is labeled +1 or -1 depends on the location of the data sample according to the hyperplane. The continuous signed distance to the hyperplane is determined and if it is above/below a certain threshold the data sample is labeled +1/-1. We are not interested in the decision process using a threshold. Instead we are interested in the continuous signed distance to the hyperplane.

The data points we use represent facial surface geometry and the purpose is to determine a continuous gender value based on the geometry. In a first instance we compute a PCA model of the facial surface shape and texture information to obtain a data reduction. Instead of using the facial surface and texture vector representations we can now use the reduced PCA model parameter descriptions as data points. In a second instance we label every facial entry according to their gender using discrete values +1 (male) or -1 (female). Then a SVM is trained, using the available LSSVM toolbox [182] for matlab^{TM}, computing the optimal separating hyperplane in the PCA face-space. Instead of using the trained SVM for classification purposes of unseen facial data samples we extract the signed distance value of every training data sample to the hyperplane, being the continuous gender values. A more positive/negative distance to the hyperplane implies a more male/female alike face being further away from the female/male class.