



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

Local graph-based probabilistic representation of object shape and appearance for model-based medical image segmentation

Promotor:
Prof. Dr. ir. P. Suetens
Prof. Dr. ir. F. Maes

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Dieter SEGHERS

November 2008



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

Local graph-based probabilistic representation of object shape and appearance for model-based medical image segmentation

Jury:

Prof. Dr. ir. J. Berlamont (voorzitter)

Prof. Dr. ir. P. Suetens (promotor)

Prof. Dr. ir. F. Maes (promotor)

Prof. Dr. ir. D. Vandermeulen

Prof. Dr. J. Verschakelen

Dr. B. van Ginneken

(University Medical Center Utrecht)

Prof. Dr. T. Cootes

(University of Manchester)

ISBN 978-90-5682-998-8

U.D.C. 681.3*I4

Wett. dep. D/2008/7515/108

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Dieter SEGHERS

November 2008

© Katholieke Universiteit Leuven – Faculteit Ingenieurswetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

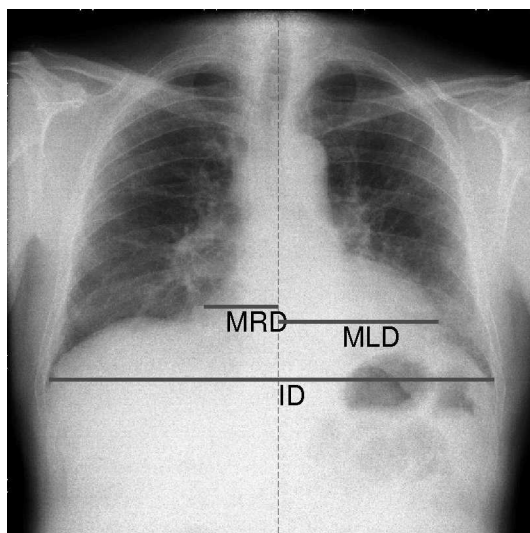
Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of this publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 978-90-5682-998-8
U.D.C. 681.3*14
Wett. dep. D/2008/7515/108

Woord vooraf

In het najaar van 2003 begon ik als wetenschappelijk medewerker van de Katholieke Universiteit Leuven aan een nieuw onderzoeksproject. Een eerste opdracht was eenvoudig, zo leek althans: “Bereken automatisch de cardiothoracale ratio uit een radiografie van de thorax”. Het was maar een eerste kleine opdracht in het ganse project. Ik hoefde maar een methode uit de literatuur toe te passen die automatisch de longen kon aflijnen en het probleem zou opgelost zijn. Echter, door (of beter gezegd dankzij) de ligging van ons labo, namelijk binnen de muren van het Universitaire Ziekenhuis Gasthuisberg, kreeg ik te maken met echte patiëntenbeelden in plaats van met ideale data zoals die in de academische wereld vaak verondersteld wordt. De meest performante methodes uit de literatuur faalden en daardoor moest ik op zoek naar een oplossing. Mijn gepassioneerde zoektocht naar een geschikte methode was de drijfveer voor dit doctoraat.



Tijdens deze zoektocht stond ik er niet alleen voor. Eerst en vooral wil ik graag mijn promotoren bedanken. Paul Suetens is het hoofd van onze groep en ondanks zijn drukke agenda kon ik op elk moment, met eender welk probleem, bij hem terecht. Paul, ik vermoed dat er maar weinigen zijn, die als hoofd van een onderzoeksafdeling zo bereikbaar zijn voor hun werknemers als jij. Bedankt dat ik mijn steentje heb mogen bijdragen in uw groep. Frederik Maes, jij hebt als echte onderzoeksprofessor met je kritische kijk mij de eerste knepen van het vak geleerd. Bedankt voor de vele wetenschappelijke discussies en bedankt voor al uw inspanningen bij het tot stand komen van dit manuscript. Verder wil ik elk jurylid uitdrukkelijk bedanken. Dirk Vandermeulen, je inzicht en literatuurkennis leverden een significante bijdrage aan dit doctoraat. De samenwerking met Johny Verschakelen was een belangrijke stimulans en zorgde ervoor dat ik me goed bewust was van de relevantie van dit werk. Het onderzoekswerk van Bram van Ginneken was een belangrijke bron van inspiratie voor mij en zijn inzet en bijdrage voor het medische beeldverwerkingsdomein zijn ronduit indrukwekkend. *Tim Cootes, your useful feedback in the few written contacts we had, have had a very large impact on this work. You referred to the relevant articles, which helped me improving my work considerably. Bram and Tim, it is a real honor to me to have you in my jury.* Tenslotte bedank ik ook nog Prof. Jean Berlamont om voorzitter te willen zijn tijdens mijn doctoraatsverdediging.

De toffe sfeer in de onderzoeksgroep werkte uiterst motiverend en zorgde ervoor dat ik telkens met plezier het ziekenhuis betrad. Ik kan niet anders dan Jeroen 'de pletskop' Hermans, vriend en ex-collega, een speciale vermelding geven. Hij was gedurende de laatste jaren degene waarmee ik de ganse werkdag zat opgescheept. De humor, de talloze wetenschappelijke discussies en de vele babbels waren heel inspirerend en maakten het onderzoeksleven zowel plezierig als productief.

Bart en Dominique, bedankt voor alle hulp bij de praktische en technische problemen. Annitta en Patricia, ik waardeer de glimlach waarmee jullie mij door alle administratieve problemen hielpen. Bedankt aan al mijn collega's op wiens steun ik heb kunnen rekenen en die voor een leuke sfeer zorgden: An Elen, Annemie Ribbens, Dirk Loeckx, Dirk Smeets, Emiliano D'Agostino, Filip Schutyser, Gabriel Kiss, Gert Merckx, Gert Peeters, Janaki Rangarajan, Jeroen Wouters, Johan Van Cleynenbreugel, Johannes Keustermans, Lennart Scheys, Liesbet Roose, Maarten Depypere, Miet Loubele, Olivier Pareit, Peter Claes, Pieter 'PB' Brunyninckx, Pieter De Groeve, Pieter Slagmolen, Qian Wang, Siddy Srivastava, Stijn De Buck, Sven De Greef, Thomas Fabry, Wouter Mollemans en Yves Lambelin.

Verder bedank ik nog een aantal mensen van buiten de groep met wie ik heb samengewerkt. *Stelios, thank you for delineating all these lung images by hand.* Ook wil ik de mensen van Agfa Gevaert (Piet Dewaele, Kris Verstreken, Gert Behiels) bedanken, dankzij hen heb ik dit project kunnen uitvoeren.

Ik kan hier onmogelijk mijn vrienden vergeten. Ze hebben er heel vaak voor gezorgd dat ik me op de gepaste manier kon ontspannen. Op de eerste plaats komen mijn *ex-roomies* van ‘de mechelse 203’. Gert, Bert, Wout en Yves, het was voor mij en Tinne een onvergetelijke tijd en die pakken ze ons niet meer af. Ik ben er zeker van dat we qua sfeer zeker niet moesten onderdoen voor die van ‘*Friends*’. Bedankt ook aan alle vrienden van de Leuven-kliek en de Fakbar-kliek voor de vele feestjes.

Ik zou niet meer kunnen tellen hoeveel avonden, weekends en vakantiedagen zijn opgeofferd aan het schrijven van deze tekst. Mijn grootste dank gaat daarom uit naar Tinne, mijn vriendin. Tinne, bedankt dat je altijd voor me klaar stond, altijd goedgehumt en bereid om te luisteren en zelfs om alleen verder te klussen in ons huis.

Tenslotte bedank ik mijn familie, op wie ik altijd kon rekenen en die wel eens bezorgd was. Bedankt papa & mama, bart, pepe & meme (“Zie je nu wel dat ik nooit gestopt was met dit doctoraat?”), opa & oma en familie van Tinne.

Dieter
Leuven, 3 november 2008.

Abstract

Local graph-based probabilistic representation of object shape and appearance for model-based medical image segmentation

Image segmentation is the process of partitioning a digital image into regions originating from different objects in the scene. The segmentation of anatomical objects is indispensable for the analysis of medical images. It enables the assessment of anatomical measurements and it is a possible means towards diagnosis, therapy planning and visualization.

As anatomical objects appear in medical images with high variability, the construction of a model that incorporates prior knowledge about these objects is essential during segmentation. A popular and very effective approach is to represent the shape as a set of landmark points, and learn the shape variations from a set of example shapes. Whereas conventional methods build a global point distribution model that considers correlations between all points in the set, this thesis presents a localized model that captures statistical prior shape information as a concatenation of multiple local shape models into a deformable graph configuration. The method has a strong theoretical basis as the model construction and model fitting are formulated from a probability point of view. Its validity and highly generic nature are illustrated for the segmentation of multiple anatomical structures, both from two- as three-dimensional images. A comparison to methods that use a global model shows that the presented approach, thanks to its localized nature, is able to fit more accurately to unseen objects.

Lokale graafgebaseerde probabilistische representatie van beeldobjecten voor modelgebaseerde segmentatie van medische beelden

Beeldsegmentatie is het proces waarbij een beeld ingedeeld wordt in regio's afkomstig van verschillende objecten. De segmentatie van medische beelden laat toe om anatomische metingen te maken, maar het kan ook de nodige stap zijn voor het stellen van een diagnose, voor het plannen van een therapie of voor de visualisatie van anatomische structuren.

Omdat anatomische objecten zich in medische beelden met een grote verscheidenheid manifesteren, is het noodzakelijk om een maximum aan voorkennis over deze objecten te gebruiken bij het segmenteren ervan. Een populaire en effectieve strategie stelt het object voor als een verzameling gemarkeerde punten, en leert de typische vormvariaties uit een verzameling voorbeelden. Conventionele methodes bouwen een globaal punt distributie model dat correlaties tussen alle punten van de vorm beschouwt. Deze thesis daarentegen, stelt een gelokaliseerd model voor, dat vormkennis beschrijft als een aanschakeling van meerdere lokale statistische modellen geconfigureerd in een graafstructuur. De methode heeft een sterke theoretische basis doordat zowel het bouwen van het model als het toepassen van het model geformuleerd zijn vanuit de waarschijnlijkheidstheorie. De kracht en het generische karakter van de segmentatiemethode worden geïllustreerd aan de hand van meerdere anatomische structuren, zowel in twee- als driedimensionale beelden. Een vergelijkende studie met methodes die een globaal model gebruiken leert dat deze strategie, dankzij zijn gelokaliseerde karakter, toelaat om ongeziene objecten nauwkeuriger te segmenteren.

Lokale graafgebaseerde probabilistische representatie van beeldobjecten voor modelgebaseerde segmentatie van medische beelden

Nederlandse samenvatting

1 Inleiding

1.1 Medische beeldvorming

Door de ontdekking van X-stralen in 1895 [1] werd de mens zich bewust van de mogelijkheid om de binnenkant van het menselijk lichaam in beeld te brengen. Deze elektromagnetische golven, ook wel eens Röntgenstralen genaamd, dringen doorheen het menselijk weefsel en verzwakken geleidelijk afhankelijk van het weefseltype. Menselijk bot bijvoorbeeld, absorbeert de stralen meer dan zacht weefsel. Radiografie maakt gebruik van deze eigenschap om een beeld te vormen van de menselijke anatomie. Een bundel X-stralen, gevormd in een Röntgenbuis, wordt door het lichaam gestuurd en de overblijvende fotonen worden opgevangen op een film of een digitale receptor. Röntgens historische experiment stond aan de wieg van de radiologie. Dit is het deel van de geneeskunde dat zich bezig houdt met de acquisitie van medische beelden voor de diagnose en behandeling van ziektes. De uitvinding van de tomografie was een tweede doorbraak. Deze techniek laat toe om een dwarse doorsnede van het lichaam in beeld te brengen. De CT-scanner (*Computed Tomography*), voor het eerst commercieel beschikbaar in de jaren 1970 [2], genereert een volumetrisch beeld van het lichaam als een reeks beelden die elk een doorsnede op een bepaalde positie van het lichaam tonen. Hierdoor is het mogelijk om met beeldvorming de ware driedimensionale vorm van anatomische structuren te bestuderen.

Vandaag de dag gebruikt de moderne beeldvorming niet enkel X-stralen, maar verschillende fysische principes [3]. Zo worden bijvoorbeeld magnetische velden

gebruikt door MR-scanners (*Magnetic Resonance*) en maakt PET-beeldvorming (*Positron Emission Tomography*) gebruik van radioactiviteit.

1.2 Segmentatie van anatomische structuren

Door de opkomst van de driedimensionale (3D) beeldmodaliteiten en de digitalisering van de radiografie, onstond een nieuw domein in de computerwetenschappen: medische beeldverwerking. Onderzoekers in dit domein ontwikkelen methodes voor de kwantitatieve analyse en visualisatie van medische beelden. Registratiemethodes zoeken het geometrische verband tussen twee of meerdere beelden en maken het mogelijk om op die manier informatie uit meerdere beelden te combineren en samen te visualiseren. Bijvoorbeeld bij het opvolgen van longkanker vergelijkt de radioloog twee CT-scans die op verschillende tijdstippen werden opgenomen. Om na te gaan of een bepaalde nodule of een tumor al dan niet gegroeid is, is het noodzakelijk dat de twee beelden gealigneerd worden vooraleer ze aan visuele inspectie te onderwerpen.

Een andere belangrijke groep algoritmes tracht een beeld te segmenteren in betekenisvolle regio's. Algemeen is beeldsegmentatie gedefinieerd als het partitioneren van een beeld in regio's die afkomstig zijn van verschillende objecten. Soms doelt de term segmentatie op het groeperen van pixels in verschillende labels, maar het kan ook duiden op het aflijnen van specifieke objecten in het beeld. Segmentatie heeft heel wat toepassingen in de radiologie. Het laat toe om anatomische metingen te maken, maar het kan ook de nodige stap zijn voor het stellen van een diagnose, voor het plannen van een therapie of voor de visualisatie van anatomische structuren.

Deze thesis behandelt de segmentatie van anatomische objecten met een consistente geometrie. Dit betekent dat anatomisch corresponderende punten kunnen gedefinieerd worden tussen verschillende instanties van hetzelfde object. De longen en de lever bijvoorbeeld, voldoen aan deze voorwaarde, in tegenstelling tot bijvoorbeeld pathologische structuren die een minder consistente vorm hebben.

Anatomische objecten vertonen typisch een heel grote variabiliteit in medische beelden. De vorm is vaak heel complex en bovendien bestaan er ook grote verschillen tussen meerdere instanties van dezelfde structuur. Ook vertoont de fotometrie van de beelden een grote variabiliteit, zowel binnen één beeld, als tussen meerdere subjecten. Dit gegeven bemoeilijkt de automatische segmentatie van anatomische structuren uit medische beelden, en maakt het noodzakelijk om een maximum aan voorkennis over deze objecten te modelleren vooraleer deze structuren te segmenteren.

Segmentatie met *active shape models* (ASM) [4], een populaire en effectieve strategie, stelt het object voor als een verzameling gemarkeerde punten (merkpunten), en leert de typische vormvariëaties uit een verzameling voorbeelden. Het vormmodel (punt distributie model) bestaat uit de gemiddelde vorm samen met de belangrijkste modes van vormvariatie. Om het object te segmenteren in een nieuw ongezien beeld, wordt een lineaire combinatie van deze modes van variatie gezocht, zodat de vorm optimaal past op de fotometrische ken-

merken van het beeld. Ondanks het succes van deze aanpak, werden in de literatuur een aantal tekortkomingen (vb., [5, 6]) blootgelegd. Het gebruik van een globaal model maakt het moeilijk om het globale optimum te vinden. De gebruikte optimalisatie-algoritmes komen vaak in een lokaal optimum terecht. Bovendien is vaak een te grote leerverzameling nodig om ervoor te zorgen dat het model nauwkeurig nieuwe ongeziene vormen kan reconstrueren. Waar conventionele methodes een globaal punt distributie model (PDM) bouwen, dat correlaties tussen alle punten van de vorm beschrijft, wordt in deze thesis een gelokaliseerd model voorgesteld. De vormkennis wordt beschreven als een aaneenschakeling van meerdere lokale statistische modellen geconfigureerd in een graafstructuur.

Het gelokaliseerde karakter van dit model zorgt voor een verhoogde flexibiliteit om nieuwe objecten nauwkeuriger te kunnen reconstrueren. Deze hypothese zal worden onderzocht aan de hand van de segmentatie van meerdere anatomische objecten, zowel uit twee- als driedimensionale beelden.

In de volgende paragrafen wordt uitgelegd hoe het model wordt opgebouwd, vervolgens hoe het model wordt gebruikt om nieuwe beelden te segmenteren en tenslotte worden een aantal concrete toepassingen beschreven.

2 Modelleren van beeldobjecten

De automatische segmentatie van anatomische structuren vereist voorkennis over enerzijds de vorm van het object, en anderzijds over de typische fotometrie van de beelden. Een beschrijving van de vorm wordt een vormmodel genoemd en het intensiteitsmodel beschrijft de voorkennis over de fotometrie.

De vorm- en intensiteitsmodellen worden geschat uit een leerverzameling. Deze verzameling bestaat uit intensiteitsbeelden \mathbf{I} met bijbehorende correct gesegmenteerde objecten \mathbf{l} . Zowel het beeld \mathbf{I} als de vorm \mathbf{l} van het anatomische object worden beschouwd als het resultaat van een toevalsproces. Segmentatie kan worden geïnterpreteerd als de oplossing van volgend *maximum-a-posteriori*-probleem:

$$\max_{\mathbf{l}} p(\mathbf{l}|\mathbf{I}) = \max_{\mathbf{l}} \frac{p(\mathbf{l})p(\mathbf{I}|\mathbf{l})}{p(\mathbf{I})} = \max_{\mathbf{l}} p(\mathbf{l})p(\mathbf{I}|\mathbf{l})$$

In woorden betekent dit: zoek de meest waarschijnlijke vorm die bij het gegeven beeld past. Het bouwen van een model concretiseert zich nu in het schatten van de kansverdelingen $p(\mathbf{l})$ (vormmodel) en $p(\mathbf{I}|\mathbf{l})$ (intensiteitsmodel) uit de leerverzameling. Dit is het onderwerp van deze paragraaf. Het oplossen van het maximalisatie probleem komt in de volgende paragraaf aan bod.

2.1 Vormmodel

Het object wordt voorgesteld als een graaf

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

waar de knopen $\mathcal{V} = \{1, \dots, n\}$ refereren naar merkpunten $\mathbf{l}_1, \dots, \mathbf{l}_n$ die op consistente plaatsen op de vorm zijn gedefinieerd. De verbindingen \mathcal{E} bevatten t ongeordende paren van knopen $\{i, j\}$. Als alternatief kan men de verbindingen beschrijven aan de hand van de omgevingsverzameling $\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{V}\}$ waaring \mathcal{N}_i de burens van knoop i bevat. Fig. 1 toont een aantal voorbeelden van anatomische objecten die worden beschreven als een graaf.

Om het vormmodel op te bouwen is het nodig om de kansverdeling $p(\mathbf{l})$ te schatten. Om onze vooropgestelde eis om meer flexibiliteit (t.o.v. globale PDMs) in te bouwen, wordt de veronderstelling gemaakt dat de vorm zich gedraagt als een Markov random veld (MRF). Dit betekent dat de ligging van een punt enkel rechtstreeks gecorreleerd is met de ligging van zijn burens. Wiskundig wordt dit uitgedrukt als:

$$p(\mathbf{l}_i \mid \mathbf{l}_{\mathcal{V}-\{i\}}) = p(\mathbf{l}_i \mid \mathcal{N}_i)$$

Als de ligging van de burens van een punt \mathbf{l}_i gekend zijn, dan levert de kennis over de ligging van alle andere (verderop gelegen) punten geen extra informatie op voor de kansverdeling van \mathbf{l}_i .

Een kansverdeling $p(\mathbf{l})$ kan in de log-ruimte worden uitgedrukt:

$$p(\mathbf{l}) = \frac{1}{Z_1} \exp\left(-\frac{1}{2}E_1(\mathbf{l})\right)$$

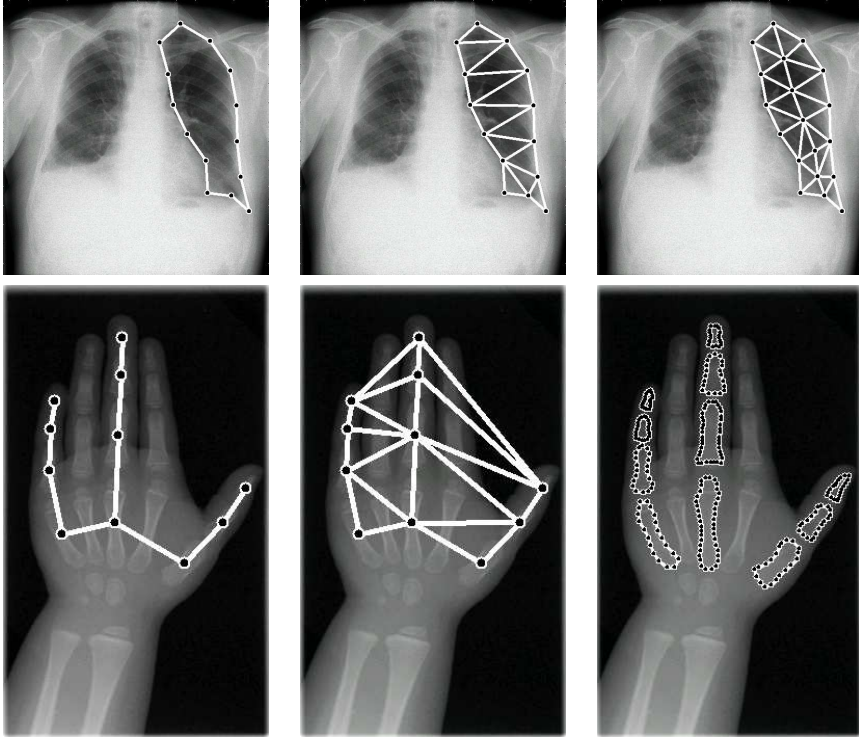
met $E_1(\mathbf{l})$ de zogenaamde vormenergie. Dit impliceert dat een vorm met een hoge energie onwaarschijnlijk is en omgekeerd, een vorm met een lage energie heeft een grote kansdichtheid.

Het Hammersley-Clifford theorema [7] drukt uit dat de energie van een Markov random veld kan geschreven worden als een som van lokale potentiaalfuncties over alle mogelijke kliks in de graaf. Een klik is een deelverzameling van de graaf, waarbij elke knoop in de deelverzameling verbonden is met alle andere knopen in die verzameling. Zo zal bijvoorbeeld de vormenergie van een contour kunnen geschreven worden als een som van lokale functies van maximaal twee variabelen en bevat de energie van een triangulatie ook derdegraadspotentiaalfuncties.

In deze thesis wordt een tweedegraadsbenadering gemaakt. Daardoor kan de vormenergie van eender welke graaf worden uitgedrukt als een som van tweedegraadsenergieën:

$$E_1(\mathbf{l}) = \frac{n-1}{2t} \sum_{i=1}^n \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$$

met $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$ de energie van de verbinding tussen de punten \mathbf{l}_i en \mathbf{l}_j . Deze drukt de waarschijnlijkheid uit van de ligging van \mathbf{l}_j ten opzichte van \mathbf{l}_i . De onbekende parameters in deze potentiaalfuncties worden geschat aan de hand van de leerverzameling.



Figuur 1: De vorm van een anatomisch object kan worden voorgesteld als een graaf.

2.2 Intensiteitsmodel

Het intensiteitsmodel $p(\mathbf{I}|\mathbf{l})$ drukt de waarschijnlijkheid uit om een beeld \mathbf{I} te observeren gegeven de ligging \mathbf{l} van het object. Bij het schatten van dit kansmodel maken we twee vereenvoudigingen.

1. Ten eerste wordt verondersteld dat de vorm \mathbf{l} enkel invloed heeft op de fotometrie \mathbf{f} in de buurt van \mathbf{l} . Deze lokale fotometrie \mathbf{f} wordt berekend door de lokale intensiteitspatronen \mathbf{f}_i rond de merkpunten \mathbf{l}_i uit het beeld te onttrekken: $\mathbf{f}_i = F(\mathbf{I}, \mathbf{l}_i)$. Dus, voor elk merkpunt wordt een kenmerkenvector uit het beeld berekend die het lokale intensiteitspatroon rond dit punt bevat.
2. Een tweede veronderstelling drukt uit dat de patronen \mathbf{f}_i van de verschillende merkpunten statistisch onafhankelijk zijn. Hierdoor kunnen we schrijven dat

$$p(\mathbf{I}|\mathbf{l}) \propto \prod_{i=1}^n p(\mathbf{f}_i)$$

met $p(\mathbf{f}_i)$ de kans dat rond het merkpunt \mathbf{l}_i de lokale beeldstructuur \mathbf{f}_i wordt waargenomen.

Het schrijven van deze kansverdeling in de log-ruimte of de enerieruimte:

$$p(\mathbf{f}_i) = \frac{1}{Z_{\mathbf{f}_i}} \exp\left(-\frac{1}{2}d_i(\mathbf{f}_i)\right)$$

laat toe om met energieën in plaats van met kansdichtheden te rekenen. De parameters in deze intensiteitspotentialen kunnen geschat worden aan de hand van de leerverzameling. Een punt \mathbf{l}_{test} in het beeld \mathbf{I} met een lage intensiteitsenergie $d_i(F(\mathbf{I}, \mathbf{l}_{\text{test}}))$ heeft een lokaal beeldpatroon dat gelijkaardig is aan de beeldpatronen in de leerverzameling rond ditzelfde merkpunt.

Het globale intensiteitsmodel kan nu geschreven worden als

$$p(\mathbf{I}|\mathbf{l}) = \frac{1}{Z_{\mathbf{I}|\mathbf{l}}} \exp\left(-\frac{1}{2}E_f(\mathbf{I}, \mathbf{l})\right)$$

met $Z_{\mathbf{I}|\mathbf{l}}$ een normalisatieconstante en $E_f(\mathbf{I}, \mathbf{l})$ de globale intensiteitsenergie

$$E_f(\mathbf{I}, \mathbf{l}) = \sum_{i=1}^n d_i(F(\mathbf{I}, \mathbf{l}_i))$$

Deze energie drukt uit hoe goed een vorm op de fotometrische kenmerken van het beeld past.

Er zijn verschillende mogelijkheden om de intensiteitsdescriptor $F(\mathbf{I}, \mathbf{l})$ te ontwerpen. In deze thesis wordt de volgende strategie gevolgd.

1. Eerst worden een aantal kenmerkenbeelden uit het originele beeld afgeleid. Als kenmerkenbeelden opteerden we voor LOIs [8, 9] (*Locally Orderless Images*). Deze worden als volgt berekend. Eerst wordt het originele beeld verwerkt door een Gaussiaanse filterbank, die afgeleiden van verschillende ordes berekent op meerdere schalen. Vervolgens worden uit elk gefilterd beeld een aantal kenmerkenbeelden bekomen door op elke pixel (voxel) in het beeld de eerste statistische momenten te berekenen van de intensiteiten binnen een lokaal Gaussiaans venster.
2. Om de beeldstructuur rond het merkpunt \mathbf{l}_i te extraheren, worden de LOIs lokaal rond \mathbf{l}_i bemonsterd met sferische profielen. Bij tweedimensionale beelden wordt het intensiteitsbeeld bemonsterd op punten van een cirkel, gecentreerd rond \mathbf{l}_i . Bij driedimensionale beelden wordt de cirkel vervangen door een bol.

3 Segmentatie

Het segmenteren van een beeld \mathbf{I} wordt vertaald naar het vinden van de meest waarschijnlijke vorm \mathbf{l}^* , gegeven het beeld \mathbf{I} :

$$\mathbf{l}^* = \arg \max_{\mathbf{l}} p(\mathbf{l})p(\mathbf{I}|\mathbf{l})$$

Dit is equivalent met het minimaliseren van de volgende energiefunctie:

$$E(\mathbf{l}, \mathbf{I}) = \sum_{i \in \mathcal{V}} d_i(F(\mathbf{I}, \mathbf{l}_i)) + \frac{n-1}{2t} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$$

De eerst sommatieterm is de intensiteitsenergie die de vorm naar de juiste fotometrische kenmerken in het beeld drijft. De tweede term is de vormenergie en controleert de vorm.

Omdat de doelfunctie niet convex is en een groot aantal onafhankelijke variabelen heeft, sturen conventionele optimalisatietechnieken de oplossing naar foute lokale minima. Om dit probleem het hoofd te bieden, wordt het probleem geconverteerd naar een discreet optimalisatieprobleem, door voor elk merkpunt een beperkt aantal geschikte kandidaten te selecteren. Dit wordt gerealiseerd door het beeld, of een relevante regio in het beeld, af te speuren met de intensiteitspotentialen $d_i(F(\mathbf{I}, \mathbf{l}_i))$. Voor elk merkpunt worden de m goedkoopste locaties geselecteerd, resulterend in de verzameling kandidaten:

$$\mathcal{C} = \{\{\mathbf{l}_{ik}, \mathbf{f}_{ik}\}_{k=1}^m\}_{i=1}^n$$

Op deze manier is het segmentatieprobleem vertaald naar een labelprobleem. Voor elke knoop in de graaf, dient één kandidaat geselecteerd te worden. Dit wordt wiskundig voorgesteld met binaire toekenningvectoren \mathbf{x}_i : $\sum_{k=1}^m x_{ik} = 1$ en $x_{ij} = 1$ betekent dat kandidaat j wordt geselecteerd in knoop i . Het minimalisatieprobleem kan nu worden geschreven als

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} E(\mathbf{x}, \mathcal{C})$$

met

$$E(\mathbf{x}, \mathcal{C}) = \sum_{i \in \mathcal{V}} \sum_{a=1}^m x_{ia} \left(d_i(\mathbf{f}_{ia}) + \frac{n-1}{2t} \sum_{j \in \mathcal{N}_i} \sum_{b=1}^m x_{jb} d_{ij}(\mathbf{l}_{ia}, \mathbf{l}_{jb}) \right)$$

De totale energie bestaat uit eerstegraads-intensiteitsenergieën toegekend aan de knopen van de graaf \mathcal{V} , en tweedegraads-vormenergieën toegekend aan de verbindingen \mathcal{E} .

In deze thesis worden verschillende optimalisatietechnieken toegepast om dit combinatorische probleem op te lossen. Eerst wordt aangetoond dat een globaal optimum kan worden gevonden door gebruik te maken van dynamisch programmeren (DP) [10]. Jammer genoeg schaalde de rekencomplexiteit van deze methode op met de complexiteit van de graafconfiguratie. Praktisch komt het erop neer dat DP enkel toegepast wordt bij objecten die worden voorgesteld als contouren. Daarom werden nog twee alternatieve technieken toegepast. *Mean field annealing* [7] maakt gebruik van het feit dat het labelprobleem \mathbf{x} zich gedraagt als een MRF ten opzichte van $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. De rekencomplexiteit van dit algoritme is onafhankelijk van de configuratie van de graaf, maar kan wel hoog oplopen door het hoge aantal benodigde iteraties. Een derde algoritme splits het probleem op in meerdere gemakkelijk op

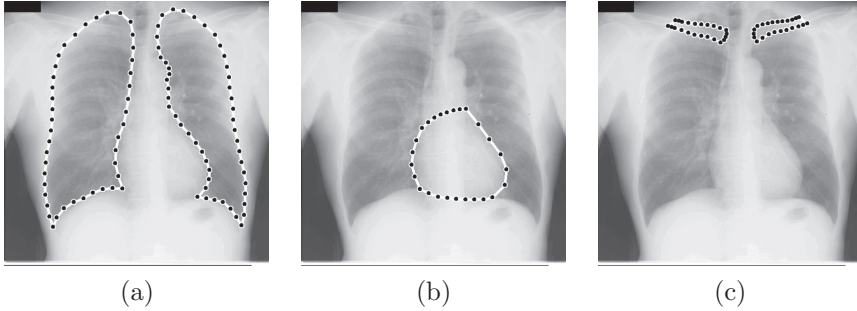
te lossen deelproblemen (met DP) door de graaf op te delen in verschillende overlappende deelgrafen. Iteratief worden dan de minst gekozen kandidaten verwijderd tot convergentie bereikt wordt.

4 Toepassingen

Het segmenteren van medische beelden is zowel een twee- als een driedimensionaal probleem. Een radiografie toont tweedimensionale projecties van driedimensionale vormen. De afijning van deze geprojecteerde vormen is vaak uitdagend vanwege de grote vormvariëaties afkomstig van zowel populatiegerelateerde verschillen als verschillen in pose. Het overlappen van andere structuren zorgt voor een bijkomende moeilijkheid. In deze thesis werden meerdere anatomische objecten gesegmenteerd uit radiografieën: longvelden, hart en sleutelbeenderen uit thoraxbeelden, femur en tibia uit beelden van de knie en handbotjes uit radiografieën van de handen. Bovendien werd ook nog de driedimensionale vorm van de lever gesegmenteerd uit CT-scans.

In deze samenvatting bespreken we de resultaten van het segmenteren van anatomische structuren uit radiografieën van de thorax. Voor deze toepassing werd een publiekelijk beschikbare databank gebruikt. De SCR-databank [11, 12] (*Segmentation in Chest Radiographs*) werd opgebouwd ter vergelijking van algoritmes voor de segmentatie van de longvelden, het hart en de sleutelbeenderen uit radiografieën van de thorax (Fig. 2). De databank bestaat uit 247 beelden (2048×2048 pixels) uit de JSRT-databank (database [13]) waarin de anatomische objecten door twee experts manueel werden afgelijnd. De segmentaties van de eerste expert worden beschouwd als referentie. De resultaten van de automatische algoritmes en de afijningen van de tweede waarnemer worden vergeleken met de referentiesegmentaties door middel van de overlapingscoëfficiënt Ω en een afstandsmaat Δ die de gemiddelde afstandsfout langs de curves berekent. Het ultieme doel voor een automatisch algoritme is het behalen van een fout die vergelijkbaar of zelfs lager is dan de fout gemaakt door de tweede waarnemer. Een eerste helft van de beelden wordt gebruikt als leerverzameling om de modellen te bouwen, die dan worden toegepast op de beelden in de tweede helft. In een tweede fase worden de rollen van de twee helften verwisseld.

De resultaten van de hier voorgestelde methode (MISCG: *Minimal Intensity and Shape Cost Graph*) en een aantal andere participerende methodes zijn samengevat in Tabel 1. Voor een overzicht en een beschrijving van alle methodes en bijhorende resultaten verwijzen we naar [11, 12]. Voor elk object heeft MISCG de hoogste performantie. De longen worden gesegmenteerd met een nauwkeurigheid die even goed is als die van manuele segmentaties. Voor het afijnen van de sleutelbeenderen en het hart lijkt de tweede expert beter te doen dan de algoritmes. Bij deze structuren is het contrast heel wat minder duidelijk dan bij de longen. Het menselijk brein heeft bij het afijnen van deze structuren minder problemen omdat het ook gebruik maakt van andere contextuele informatie zoals bijvoorbeeld andere overlappende of naburige structuren. De



Figuur 2: Anatomische structuren te segmenteren uit een radiografie: longen (a), hart (b) en sleutelbeenderen (c).

computer algoritmes gebruiken enkel voorkennis over één specifiek object. Verder wordt ook nog de pixelfout gerapporteerd. Deze foutmaat wordt berekend als de fractie van alle pixels die fout worden geclassificeerd. MISCG heeft de laagste gemiddelde fout (0.033) op het manuele resultaat (0.029) na. Pixel-classificatie haalt een score van 0.043. Methodes die gebruik maken van een globale PDM zoals *ASM tuned* en *AAM whiskers BFGS* scoren minder goed met een fout van respectievelijk 0.044 en 0.046. De belangrijkste verklaring hiervoor is het gebrek aan flexibiliteit van de globale PDMs. Bovendien heeft MISCG het voordeel dat een globaal optimum wordt gevonden, in tegenstelling tot de globale PDMs.

5 Conclusies

In deze thesis werd een modelgebaseerde methode voor de segmentatie van medische beelden voorgesteld, die in staat is om ongeziene beelden nauwkeurig te segmenteren, zelfs indien maar een beperkte leerverzameling beschikbaar is. Waar de meeste algoritmes gebruik maken van een globale PDM, aangeleerd uit een aantal trainingsbeelden met bijhorende referentiesegmentaties werd hier een model voorgesteld dat enkel lokale correlaties in rekening brengt. Deze aanpak heeft twee grote voordelen. Ten eerste, het model kan ongeziene objecten nauwkeurig reconstrueren, zelfs indien maar een beperkt aantal voorbeelden beschikbaar is. Ten tweede, het segmentatieprobleem is omgezet in een combinatorisch optimalisatieprobleem waarvan een optimale of bijna-optimale oplossing kan gevonden worden met discrete technieken als dynamisch programmeren en *mean field annealing*.

De methode heeft een sterk gefundeerde basis dankzij een formulering vanuit de waarschijnlijkheidstheorie. De intensiteiten van het beeld en de vorm van het object worden beschouwd als gecorreleerde toevalsprocessen. Deze strategie laat toe om het segmentatieprobleem te formuleren als een *maximum-a-posteriori*-probleem: gegeven een beeld, vind de meest waarschijnlijke vorm

	$\mu \pm \sigma$	Min	Q1	Median	Q3	Max
<i>Longen</i>						
MISCG	95.3 ± 02.6	83.4	94.7	95.8	96.4	97.7
PC post-processed	94.5 ± 02.2	82.3	93.9	95.1	95.8	97.2
Tweede waarnemer	94.6 ± 01.8	82.2	93.9	94.9	95.8	97.2
ASM tuned	92.7 ± 03.2	74.5	91.7	93.6	94.6	94.6
AAM whiskers BFGS	92.2 ± 02.9	71.8	91.4	93.1	94.0	96.1
<i>Hart</i>						
Tweede waarnemer	87.8 ± 05.4	57.1	84.3	88.8	91.6	96.5
MISCG	84.9 ± 10.4	26.1	82.4	87.4	90.7	96.3
AAM whiskers BFGS	83.4 ± 07.0	51.0	79.1	84.5	88.2	96.7
PC post-processed	82.4 ± 07.7	50.0	78.3	84.4	87.7	93.2
ASM tuned	81.4 ± 07.6	52.0	77.0	82.7	87.3	93.8
<i>Sleutelbeenderen</i>						
Tweede waarnemer	89.6 ± 03.7	70.7	88.0	90.5	92.2	95.2
MISCG	77.5 ± 12.9	01.1	73.8	80.7	85.4	91.5
ASM tuned	73.4 ± 13.7	09.3	70.5	77.6	82.2	91.2
AAM whiskers BFGS	64.2 ± 17.1	00.3	58.8	68.9	76.1	86.1
PC post-processed	61.5 ± 12.3	22.3	55.4	63.9	70.6	83.7

Tabel 1: Overlappingscoefficient (%) voor elke anatomische structuur. Het minimum, maximum, de mediaan en het eerste en derde kwartiel worden gerapporteerd. De resultaten zijn gerangschikt volgens de mediaan.

van het object. Een belangrijk voordeel van deze aanpak is dat een zinvolle afweging gemaakt wordt tussen geometrische en fotometrische karakteristieken.

Het performante en uiterst generische karakter van de methode werd aangetoond door het succesvol segmenteren van meerdere anatomische objecten, zowel in twee- als driedimensionale beelden.

List of acronyms and symbols

List of acronyms

2D 2-dimensional

3D 3-dimensional

AAM active appearance model

ASM active shape model

ASMOF active shape model with optimal features

CT computed tomography

CTR cardiothoracic ratio

DP dynamic programming

DPCC dynamic programming closed contour

GRF Gibbs random field

ICP iterative closest point

IDP iterative dynamic programming

JSRT Japanese Society of Radiological Technology

kNN k-nearest neighbour

LOI locally orderly images

MAP maximum a posteriori

MCP minimal cost path

MDL minimal description length

MFA	mean field annealing
MISCG	minimal intensity and shape cost graph
MR	magnetic resonance
MRI	magnetic resonance imaging
MRF	Markov random field
PA	posterior-anterior
PC	pixel classification
PCA	principal components analysis
PET	positron emission tomography
PDM	point distribution model
PDF	probability density function
ROI	region of interest
RMS	root mean square
SCR	segmentation in chest radiographs
SSD	sum of squared differences

List of symbols

$\langle \mathbf{x} \rangle$	Expected value of \mathbf{x}
a	Search grid extent
\mathcal{C}	Candidate set $\mathcal{C} = \{\{\mathbf{l}_{ik}, \mathbf{f}_{ik}\}_{k=1}^m\}_{i=1}^n$
$d_i(\mathbf{f}_i)$	Intensity energy of \mathbf{f}_i
$d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$	Edge energy of $\mathbf{l}_j - \mathbf{l}_i$
$d_{ik ij}(\mathbf{l}_i, \mathbf{l}_j, \mathbf{l}_k)$	Conditional edge energy of $\mathbf{l}_k - \mathbf{l}_i$ given $\mathbf{l}_j - \mathbf{l}_i$
\mathbf{e}_{ij}	Edge vector: $\mathbf{e}_{ij} = \mathbf{l}_j - \mathbf{l}_i$
$E_1(\mathbf{l})$	Shape energy of \mathbf{l}
$E_{\mathbf{f}}(\mathbf{I}, \mathbf{l})$	Intensity energy of \mathbf{l} in image \mathbf{I}
$E(\mathbf{I}, \mathbf{l})$	Object energy of \mathbf{l} in image \mathbf{I}
$E(\mathbf{x}, \mathcal{C})$	Discrete object energy of labeling \mathbf{x} of \mathcal{C}
E^*	Minimal value of $E(\mathbf{I}, \mathbf{l})$

E_i	Remaining energy after eliminating vertex i
\mathcal{E}	Edges of the graph: $\mathcal{E} = \{\{i, j\} i, j \in \mathcal{V}\}$
f_{red}	IDP candidate reduction
\mathbf{f}_i	Intensity feature vector at vertex i
$\mathbf{f}_i^{[j]}$	Intensity profile of feature j and vertex i
\mathbf{f}_{ik}	Intensity feature vector of candidate k for vertex i
$F(\mathbf{I}, \mathbf{l})$	Gray-level pattern in image \mathbf{I} at location \mathbf{l}
\mathbf{I}	Image
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	Graph consisting of vertices \mathcal{V} and \mathcal{E}
\mathbf{l}	Shape consisting of landmark points: $\mathbf{l} = \{\mathbf{l}_1, \dots, \mathbf{l}_n\}$
\mathbf{l}_i	Landmark point at vertex i
\mathbf{l}_{ik}	Location of candidate k for vertex i
\mathbf{l}^*	Shape that minimizes $E(\mathbf{I}, \mathbf{l})$
L	LOI derivative operator
m	Number of candidates
M	LOI statistical moment index
n	Number of landmarks
n_d	Number of image dimensions
$n_{\mathbf{f}}$	Number of samples in spherical profile
$n_{\mathbf{I}}$	Number of LOIs
n_T	Number of annealing temperatures
\mathcal{N}	Neighborhood system of the graph: $\mathcal{N} = \{\mathcal{N}_i \forall i \in \mathcal{V}\}$
\mathcal{N}_i	Set of vertices neighboring vertex i
p, P	Probability density and discrete probability
Q_i	Elimination energy until and including vertex i
$q_{\mathcal{A}_{i,k}}$	Subfunction elimination energy Q_i
$r_{\mathbf{f}}$	Radius of spherical profile
R_i	Optimal label of vertex i
s	Number of training images
t	Number of edges

T	Gibbs temperature
\mathcal{V}	Vertices of the graph: $\mathcal{V} = \{1, \dots, n\}$
\mathbf{x}	Labeling for all vertices: $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$
\mathbf{x}_i	Assignment vector at vertex i
\mathbf{x}^*	Optimal labeling that minimizes $E(\mathbf{x}, \mathcal{C})$
Z	Normalizing constant of Gibbs distribution
α	LOI outer scale
β	Estimates of mean field $\langle \mathbf{x} \mathcal{C} \rangle$
δ	Search grid spacing
μ	Expected value
ρ	Correction for overestimated intensity energy
σ	LOI inner scale
Σ	Covariance matrix
Φ	Eigenvectors
Λ	Closest distance error
Δ	Corresponding distance error
Ω	Overlap coefficient

Contents

Acknowledgements	i
Abstract	v
Nederlandse samenvatting	vii
1 Inleiding	vii
1.1 Medische beeldvorming	vii
1.2 Segmentatie van anatomische structuren	viii
2 Modelleren van beeldobjecten	ix
2.1 Vormmodel	ix
2.2 Intensiteitsmodel	xi
3 Segmentatie	xii
4 Toepassingen	xiv
5 Conclusies	xv
List of acronyms and symbols	xvii
1 General introduction	1
1.1 Segmentation of anatomical objects	1
1.2 Purpose and contributions	2
1.3 Overview	3
2 Segmentation strategies	5
2.1 Introduction	5
2.2 Low-level segmentation	6
2.2.1 Region-based	6
2.2.2 Edge-based	7
2.2.3 Rule-based schemes	7
2.2.4 Manual interaction	8
2.3 Model-based segmentation	9
2.3.1 Intensity-based	9
2.3.2 Shape-based	11
2.4 Related methods	13
2.4.1 Active shape models	13

2.4.2	Active appearance models	16
2.4.3	Pictorial structures	17
2.5	Motivation	18
2.5.1	Landmarks	18
2.5.2	Problems with global PDMs	19
2.5.3	Local graphical templates	20
2.6	Conclusion	22
3	Modeling image objects	23
3.1	Introduction	23
3.2	Object representation	25
3.3	Training set	27
3.3.1	Ground-truth segmentations	27
3.3.2	Correspondences	28
3.3.3	Global alignment	29
3.4	Energy space	29
3.5	Shape model	29
3.5.1	Markov random fields	30
3.5.2	Shape energy	31
3.5.3	Second order approximation	34
3.5.4	Learning from examples	36
3.6	Intensity model	38
3.6.1	Model assumptions	38
3.6.2	Intensity descriptor	39
3.6.3	Intensity energy	41
3.6.4	Learning from examples	42
3.7	Combined model	44
3.8	Conclusion	46
4	Model fitting	49
4.1	Objective function	49
4.1.1	Continuous optimization problem	49
4.1.2	Discrete optimization problem	50
4.2	Optimization techniques	51
4.2.1	Dynamic programming	51
4.2.2	Mean field annealing	57
4.2.3	Heuristic search method	59
4.3	Implementation issues	60
4.3.1	Search regions	60
4.3.2	Energy balance	61
4.4	Conclusion	63

5	Validation	67
5.1	Lung field data	68
5.2	Object representation	68
5.3	Optimization algorithm	70
5.4	Energy balance	74
5.5	Intensity descriptor	80
5.5.1	LOI parameters	80
5.5.2	Profile length	81
5.5.3	Profile radius	82
5.6	Global versus local	84
5.7	Conclusion	89
6	Applications	93
6.1	Introduction	93
6.2	Anatomical structures in chest radiographs	93
6.2.1	Data	94
6.2.2	Methods	94
6.2.3	Results	96
6.3	Cardiothoracic ratio	96
6.4	Hand bones	99
6.4.1	Data	99
6.4.2	Model construction	100
6.4.3	Model fitting	100
6.4.4	Results	101
6.5	Knee segmentation	103
6.5.1	Data	103
6.5.2	Experiments	104
6.5.3	Results	104
6.6	Liver segmentation	104
6.6.1	Data	105
6.6.2	Model construction	106
6.6.3	Model fitting	109
6.6.4	Evaluation	109
6.6.5	Results	110
7	Conclusions and future work	113
7.1	Main contributions	113
7.2	Method	113
7.3	Clinical relevance	116
7.4	Suggestions for future work	116
7.4.1	Curse of number of candidates	116
7.4.2	Optimization algorithms	117
7.4.3	Merging segmentation and registration	118
A	Conditional edge energy	121

B Expected value of the Mahalanobis distance	123
Bibliography	125
List of publications	133

Chapter 1

General introduction

1.1 Segmentation of anatomical objects

The discovery of X-rays in 1895 [1] for the first time revealed the ability to look inside the human body without perforating the skin. X-rays, also called Röntgen waves, have the property that they can travel through the body and are attenuated depending on the penetrated tissue. Bone for example attenuates X-rays much more than soft tissue does. This property is exploited to create a picture of the human anatomy by sending a beam of X-rays through the body and capturing the transmitted photons. This is the basic principle of radiography.

Röntgen's historical experiment is considered as the cradle of radiology, generally defined as the medical speciality that deals with the acquisition of medical images towards diagnosis and treatment of diseases. A second breakthrough was the invention of tomography. This technique enables to visualize a single cross section of the human body using X-rays. A computed tomography (CT) scanner, first commercially available in the early 1970's [2], uses this technique to generate a whole series of cross sectional radiographs. Hence, a volumetric image is produced that contains information about the true three-dimensional (3D) shapes of anatomical structures.

Nowadays, modern radiologic imaging is no longer limited to X-rays, and uses many other physical principles [3]. For example, magnetic fields are used in magnetic resonance imaging (MRI), radioactivity enables the acquisition of positron emission tomography (PET) scans and high frequency sound waves are used in ultrasonic imaging.

Together with the onset of the 3D imaging modalities and the digitization of radiography, a new domain in computer science, medical image analysis, was born. Researchers in image processing started to develop methods for the quantitative analysis and visualization of medical images.

Registration methods combine information from multiple images by warping one image onto another image in an anatomically meaningful way. For example during lung cancer follow-up, the radiologist examines two thorax CT

scans acquired at different time points. Evaluating whether a nodule or a lung tumor has grown is greatly facilitated if the images are aligned prior to visual inspection.

Many computer-aided detection methods intend to automatically find and classify pathological abnormalities. The automated identification and localization of lung nodules in a CT scan of the thorax (e.g., [14]) saves the radiologist expensive time. Emphysema, an obstructive lung disease often caused by exposure to tobacco smoke, can be diagnosed automatically on a CT image by running texture classification methods (e.g., [15]).

An important group of methods deals with image segmentation. In general, segmentation is defined as the process of partitioning a digital image into regions originating from different objects in the scene. The term segmentation might refer to the process of grouping and labeling pixels, but also refers to the delineation of specific objects in the image. The segmentation of anatomical objects is indispensable for the analysis of medical images. It enables the assessment of anatomical measurements and it is a possible means towards diagnosis, therapy planning and visualization. For example the delineation of the lung fields from a chest radiograph allows to measure the relative heart size and indirectly enables the diagnosis of cardiomegaly (e.g., [16]). Brain segmentation can be used to automatically find multiple sclerosis lesions in an MR image (e.g., [17]). Liver segmentation is the basis for computerized planning of liver surgery as tumor resection and donor transplantation (e.g., [18]). This thesis deals with the segmentation of anatomical objects with a consistent geometry. This means that an anatomically meaningful one-to-one mapping between different shape instances exists. The lungs and the liver satisfy this condition, contrary to pathological structures that have a rather random topology.

1.2 Purpose and contributions

Anatomical objects appear in medical images with high variability. The shape is often very complex and also large geometrical differences exist between different instances of the same structure. Also the object's image photometry shows large variability. Large photometrical differences occur inside one subject and between multiple subjects. This complicates the segmentation of these structures and consequently, prior knowledge about the object to be segmented needs to be taken into account as much as possible.

Segmentation using active shape models (ASM), put forward by Cootes *et al.* [4], extracts prior shape knowledge by learning from a set of example images with corresponding ground-truth segmentations. The shape model consists of an average shape together with the most important modes of variation observed in the training set. Very soon this approach became very popular in the medical domain (e.g., [19, 20]) and is currently, thanks to a number of improvements to the original scheme (e.g., [5, 8, 21, 22]) still state-of-the-art. However, despite its success, the ASM segmentation scheme has some limi-

tations. Its use of a global shape model complicates the search for a global optimum and a sufficiently large number of training images is needed to make the model flexible and accurate enough.

In computer vision, a class of object recognition methods [23–26] exists that uses a locally deformable graph. These methods are, thanks to their localized nature, more flexible and moreover they often allow to find a global optimum. Unfortunately, they are not generally applicable for the segmentation of anatomical structures. For example the pictorial structures algorithm described in [26] uses tree-like object representations and is therefore not capable of segmenting surfaces from volumetric images. In computer vision, these methods are considered as a solution for object recognition problems where specific features in pictures are detected, rather than as a method for segmentation. Moreover, these localized methods are rather unknown in the medical image analysis domain. Many researchers believe that a global shape model is a requirement towards robust segmentation of anatomical structures. With good reason one can indeed wonder whether using only local statistical knowledge suffices.

The main purpose of this thesis is to present such a localized segmentation scheme that captures statistical prior shape information as a concatenation of multiple local shape models into a deformable graph configuration. Its validity for the extraction of both two- as three-dimensional anatomical objects will be investigated for several applications.

1.3 Overview

Chapter 2 gives an overview of the different segmentation strategies with emphasis on their applicability to medical data. Also the proposed method is situated and motivated in the broad image segmentation field.

The algorithm consists of two phases. In the training phase, described in Chapter 3, statistical models of shape and image appearance are learnt from a training set. Chapter 4 describes how the model is fitted to an unseen image to find the shape automatically. Firstly, the model fitting is translated into a discrete optimization problem and secondly, a number of appropriate optimization techniques is discussed.

The method and its methodological choices are validated in Chapter 5. In Chapter 6, the method is applied for the segmentation of several anatomical structures from both two- and three-dimensional images. Lung fields, clavicles, heart are delineated from chest radiographs. Also hand bones and knee bones are extracted from radiographs and finally, the 3D shape of the liver is automatically segmented from contrast enhanced CT scans.

Chapter 7 draws some important conclusions related to anatomical object segmentation.

Chapter 2

Segmentation strategies

2.1 Introduction

In the literature of computer vision and medical image analysis, a true wealth of image segmentation methods is available. Because the choice for an appropriate method for a specific application is of crucial importance, this wealth might feel like a poisoned gift. As in this thesis a new segmentation scheme is presented, it is important to situate and motivate the method in the broad field of image segmentation.

A method that can handle every application does not exist for several reasons. First, the geometrical structure of the objects to be segmented varies from very simple (Fig. 2.1(a)), to very complex (Fig. 2.1(b)). Also the large differences in the image appearance of the objects initiated the use of very different strategies. Objects that appear with low contrast and without a homogeneous intensity distribution will require a more sophisticated approach than classifying the pixels according to their intensity value only. The choice for an algorithm also depends on the problem setup. Real-time applications exclude the use of several advanced model-based approaches because of their computational complexity. Also, some methods are intended to be generally applicable and do not make strong assumptions about the scene whilst others are able to incorporate prior knowledge. Obviously, many strategies are needed for what is called image segmentation.

This chapter intends to give a short overview of the different segmentation strategies with emphasis on their ability to handle medical image data. The methods are classified according to the amount of prior knowledge about the scene to be segmented. Low-level strategies, discussed in Section 2.2 typically use only very limited prior knowledge. An overview of model-based methods is given in Section 2.3. The novel segmentation method that will be proposed in Chapters 3 and 4 is situated in the field of image segmentation in Section 2.5. The most related methods are described in Section 2.4.

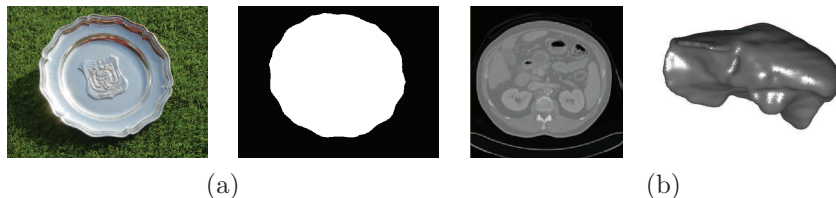


Figure 2.1: The geometrical structure of the objects to be segmented varies from very simple, as for example the object in picture (a), to very complex, as for example a liver in a CT image (b).

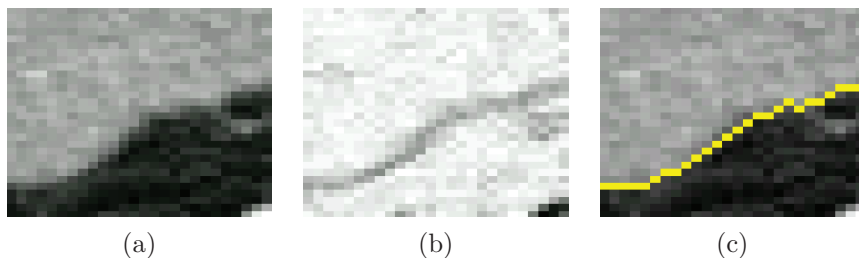


Figure 2.2: Retrieving the boundary between object and background is formulated as a minimal cost path (MCP) problem that can be solved using dynamic programming. The path that runs from the left to the right side of the image (a) is found as follows. First, the magnitude of the gradient in each pixel is computed (b). The border is then computed as a minimal cost path (c): the path is only allowed to shift one row between two consecutive nodes and the cost in the nodes is equal to the negative magnitude of the gradient.

2.2 Low-level segmentation

Low-level algorithms use very limited prior knowledge and rely on finding similarities or/and dissimilarities between neighboring pixels. Methods that group the pixels by looking at similarities are labeled as *region-based* whereas the algorithms that search dissimilarities are called *edge-based*.

2.2.1 Region-based

In this category is region growing [3] the most popular strategy. The image is partitioned into regions as follows. First, a number of seed points, one for each object, are placed in the image. Then an iterative growing procedure is started. First, all pixels adjacent to a region are compared to the respective region by computing the difference between the pixel's intensity value and the region's mean intensity value. The pixel with the smallest difference is added to the corresponding region. The algorithm stops if all pixels have been assigned to a region.

Clustering methods partition the image by searching for clusters in the feature

space of the image data. The K -means clustering algorithm [27] is a well known example. The algorithm is started from K cluster centers. Each pixel is assigned to the cluster that minimizes the difference between the feature value at that pixel and the cluster center. This procedure is repeated until convergence is obtained. The difference between a pixel and a cluster center is usually based on a combination of several features as color, intensity, texture and location.

2.2.2 Edge-based

One way to detect object boundaries in an image is to use local edge filters. These filters amplify the image pixels where a large intensity gradient is present. For example the Canny algorithm [28] is based on this technique. Many graph-based methods that search for region boundaries were proposed in the literature. In general, the image is considered as a graph where the nodes are the pixels and the edges express the connections between neighboring pixels. Montanari [29] and Martelli [30] formulated the search for the boundary between object and background as a MCP problem that can be solved using dynamic programming. The technique is demonstrated in Fig 2.2: a boundary that runs from the left to the right side of the image (Fig 2.2(a)) is found as follows. First, the magnitude of the gradient in each pixel is computed (Fig 2.2(b)). The boundary is then retrieved as the minimal cost path from left to right (Fig 2.2(c)). The path is only allowed to shift one row between two consecutive nodes and the cost in the nodes is equal to the negative magnitude of the gradient.

Other graph-based methods (e.g. [31]) assign a weight to each edge to express the dissimilarity between the neighboring pixels. A segmentation is then considered as a partitioning of the graph into subsets, such that edges in one subset have low weights compared to edges between nodes of different sets.

Wu *et al.* [32] and Shi *et al.* [33] proposed to divide the image in regions by finding minimal cuts in the graph. This approach was also applied for the extraction of surface borders from volumetric datasets [34, 35].

2.2.3 Rule-based schemes

As these low-level techniques require only limited prior knowledge about the scene to be segmented, they are generic in nature. However, for many applications they fail because of a number of reasons. The objects and regions to be partitioned are not always clearly visible due to noisy image data or if the different objects have overlapping intensity ranges. As a result, the object boundary does not necessarily result in a clearly visible edge in the image. This is often the case in medical data. For example edge detectors will fail on the fluoroscopic image shown in Fig. 2.3(a) to extract the boundaries of the vertebrae. Another problem is encountered when the objects do not have a homogeneous intensity distribution. The lung fields in the radiographic image of the chest shown in Fig. 2.3(b) exhibit a large gray-level variability due to

projections of the ribs and the clavicles. Another situation where low-level techniques fail is if the object of interest is occluded by an other object or in the case of pathologies (Fig. 2.3(c)). It is obvious that in these cases, prior knowledge about the scene to be segmented has to be incorporated by the segmentation method itself.

To overcome these obstacles many researchers developed so-called *rule-based* algorithms. Low-level techniques are used to find local relevant structures in the image and a set of rules is implemented to put the primitive structures together. Unfortunately, these methods are often heuristic and rather dedicated to handle one single problem. A new set of rules and a new implementation is required for a new application.

Despite these limitations, rule-based algorithms are still popular in commercial software because they are easy to implement and computation time is often very small. A typical example of such a rule-based approach to segment the lung fields in a chest radiograph can be found in [36].

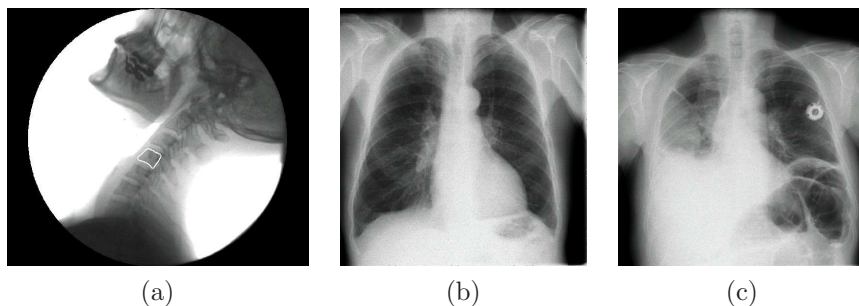


Figure 2.3: Low-level segmentation techniques often fail for medical applications. The edges of the vertebrae in the fluoroscopic image (a) are not clearly visible due to low contrast. Due to the projections of the ribs and the clavicles in the thorax radiograph (b), the intensities in the lung fields do not have a homogeneous intensity distribution. Low-level techniques also fail for the detection of the lung fields in the radiograph shown in (c) because of the occluding implant (portacath) and because of the pathology.

2.2.4 Manual interaction

Alternatively, instead of implementing a set of rules, low-level methods can also be combined with manual interaction to guide the segmentation process. Livewire [37] is such an interactive segmentation tool where the user marks points along an object boundary and the computer connects these points using dynamic programming.

2.3 Model-based segmentation

Model-based methods incorporate more extended prior information about the object to be segmented. In general, the model contains knowledge about the shape of the object and/or about the object's photometry. Many model-based methods use an explicit description of the shape and fit this shape template to the image data. An overview of these shape-based methods is reported in Section 2.3.2. Model-based techniques without an explicit shape description are discussed in Section 2.3.1.

2.3.1 Intensity-based

What differentiates these methods from the shape-based ones is that they do not fit an explicit shape model to the image data. They might include shape information but only implicitly such as an atlas. Pixel classification (PC) methods separate the pixels into classes according to photometry features. Atlas-based segmentation techniques solve the segmentation task by warping the photometry of a template image onto the photometry of the target image.

Pixel classification

PC methods consider the objects to be segmented as sets of discrete pixels rather than as objects. The pixels are labeled into classes according to their photometric features. Consequently, these methods are optimally suited for tissue classification provided that each tissue type has its own discriminating photometry. Applications where there is no well-defined relationship between pixel photometry and the corresponding labels, which is often the case when segmenting anatomical structures, need extra spatial information. This is obtained by incorporating pixel location into the feature vector.

Thresholding is the most trivial form of PC: a pixel is classified as object or background depending on whether it exceeds a preset threshold value. Thresholding is rather classified as low-level segmentation because the only prior knowledge captured is one (or multiple) preset threshold values. More advanced pixel classification methods have been published in the medical domain (e.g. [11, 36, 38–41]). They differ in which features and which classifier they use. Typical photometry features are gray-level value, entropy measures and several filter outputs. Also pixel location is often taken into account. Popular classifiers are discriminant analysis [38, 40], neural nets [38, 39], kNN-classifiers [11, 36, 40] and support vector machines. Contextual information was included by Vittitoe *et al.* [41] and Van Leemput *et al.* [42]. They used Markov random fields (MRFs) [7] to model interaction between neighboring pixels. This way, the classification of a pixel does not only depend on the features of that pixel but also on the classification of the neighboring pixels.

The pixel classifiers are learnt in a supervised or unsupervised manner. Supervised segmentation means that the model is learnt from examples. Unsupervised classification indicates that the classifier learns from the data in the

target image. In general, these approaches start from an initial model and iteratively update the model parameters (e.g. [40–42]).

Although some of these methods model shape information implicitly, either by incorporating the location of the pixel into the feature vector, or by modeling pixel interaction, the drawback of these methods is that they are unable to include knowledge about the typical shape variations. As a result, such an approach can generate object shapes that are not very meaningful. Often a postprocessing step (e.g. [11]) is used to remove isolated incorrectly classified pixels.

Atlas-based segmentation

Atlas-based methods tackle the segmentation problem by fitting an atlas image to the target image. The process of matching one image onto another image in an anatomical meaningful way is known as image registration [3]. Therefore, the atlas-based approaches are also referred to as segmentation-by-registration methods. An atlas image is a typical image of the scene to be segmented (intensity image) augmented with a corresponding golden truth segmentation (label image). The label image contains one label for each structure of interest. Once the atlas has been registered to the target, a transformation field is obtained which is subsequently applied to the atlas labels. If the matching is correct, the deformed label image indicates where each object can be found in the target image.

Atlas-based segmentation has been applied for the segmentation of structures in the brain (e.g. [43–45]). Research in this domain focusses on the registration algorithms and to atlas construction. The most simple atlas consists of a single example image and a corresponding ground-truth segmentation. Unfortunately, for many medical applications a single image atlas does not capture all the existing biological morphology seen in a population. Therefore, the construction of probabilistic atlases has received a considerable amount of attention in the medical imaging literature (e.g. [46–49]). Such an atlas contains probabilities to find specific labels at a given location in the image and is constructed from a set of example images with corresponding manual segmentations. Alternatively, instead of registering the probabilistic atlas, one can also register each individual training image to the target. Hence, each registered training image will assign a label to each pixel of the target image. A combined solution is then obtained by selecting the label that was mostly voted. Van Rikxoort *et al.* applied this approach for the segmentation of the brain caudate [45].

Atlas images have also been combined together with other segmentation techniques. For example Ali *et al.* [50] used an atlas constructed from a set of manually labeled kidney images to incorporate shape knowledge into a graph-cuts algorithm for kidney segmentation from magnetic resonance (MR) images. Van Leemput *et al.* [42] used a digital brain atlas for initialization of the classification of brain tissue.

2.3.2 Shape-based

A shape template is a mathematical representation of a shape, possibly augmented with information about the allowed geometrical variations. An image is then segmented by fitting the shape template to the image data. In general, an objective function derived from the image data provides the driving force that attracts the shape to the correct image features whereas the shape template controls the optimization process such that only plausible shapes are generated.

Methods differ in the way the shape is represented (e.g. landmark points, spline representations, implicit functions,...) and the type of information stored. Some methods use statistical distributions to describe the prior knowledge while others store information into a mathematical or a bio-mechanical model.

Rigid/affine templates

A rigid shape template is only allowed to be translated and rotated. Affine templates also allow for scaling. For example an outline of a building in a satellite image can be segmented by matching a rectangle to the image data. The rectangle is rotated, translated and resized to make it fit to the building in the image.

Hand crafted templates

A number of published methods work with hand-crafted shape templates. A famous example is the work of Yuille *et al.* [51]. They use a template consisting of a circle and parabolas to extract the eye from a picture of the face. These hand crafted algorithms have the drawback that they are limited to one single application.

Snakes and level sets

Kass and Witkin [52] proposed a more generic method, commonly known as active contour models or *snakes*, where the object contour is represented as a spline. A related method, called level set segmentation was proposed by Sethian [53]. They use implicit functions to describe the location of an object in the image. The contour (or surface) of the object is defined as its zero level set, i.e. as the points in space where the implicit function evaluates to zero.

Snakes and level sets obtain a segmentation by minimizing an energy function guided by internal forces that put constraints on the shape and external image forces that pull the object towards edges in the image. Hence, geometrical information is modeled by the internal forces and the external forces embody appearance information.

Despite their generic nature, snakes and level sets fail for many medical segmentation tasks. A first reason is that their object representation is not designed to model the typical shape variations observed among a population. Their shape model guarantees that the shape is a smooth contour or

curve rather than a realistic shape instance. Secondly, the object boundaries should coincide with edges in the image which is an often unsatisfied condition (Fig. 2.3(a)). As a result, an active contour (or surface) model will escape at an object boundary where no contrast is available. It is obvious that in these cases prior knowledge about the allowed shapes instances is needed.

Active shape and appearance models

A popular approach that satisfies this need was introduced by Cootes and Taylor [54]. They represented an object as a set of landmark points. Shape information is captured by constructing point distribution models (PDMs) using principal components analysis (PCA) of the landmarks of the shape instances in a training set after appropriate spatial normalization. By augmenting the PDMs with intensity information, these can be applied for image segmentation in the context of so-called *active shape models* (ASM) [4, 55] and *active appearance models* (AAM) [56–58].

ASM segmentation schemes augment the PDM with local landmark-individual gray-level models whereas AAM algorithms consider all object pixels to build a texture PCA model. The general AAM and ASM schemes were compared by Cootes *et al.* [59] for the alignment of structures in MR brain sections and by van Ginneken *et al.* [11] for the segmentation of the lung fields, the heart and the clavicles in posterior-anterior (PA) chest radiographs. These studies concluded that ASM performed better than AAM due to its larger capture range. However, this conclusion is out of date as some important improvements have been made to the general AAM scheme mainly focused at increasing robustness [60–62].

The active shape and appearance models are a popular strategy for the segmentation of anatomical structures (e.g. [19, 20, 63, 64]). Their success stimulated the research community to develop a number of variants to make some improvements with respect to the original scheme of Cootes and Taylor. An ASM with an improved gray-level appearance model was introduced by van Ginneken *et al.*, called ASM *with optimal features* (ASMOF) [8]. Other examples of modified ASM schemes can be found in [21, 65–67].

In general these algorithms consist of the following components: (1) a global shape model to ensure that only plausible shapes are generated, (2) an intensity model to drive the object to a region in the image with gray-level appearance similar to what is found in the training set and (3) a search method to match the model to a new image.

Graphical templates

Another series of landmark-based approaches for object detection use localized shape constraints instead of one global shape model. The shape is considered as a graph where the landmarks are the vertices in the graph and the edges define the local shape dependencies.

Fischler and Elschlager [23] introduced a pictorial structure representation

whereby an object is modeled as a collection of parts linked together through spring-like connections. A similar approach was proposed by Felzenszwalb and Huttenlocher [26, 68]. They use a tree-like representation to recognize objects by optimizing an objective function that consists of match costs for the parts (landmarks) and deformation costs for connected pairs of parts. The global optimum is found using dynamic programming (DP) [10]. This method has been applied for locating facial features in pictures [68, 69]. Similar work was published by Amit and Kong [24] and Coughlan *et al.* [25].

Overall, these methods find a global solution for the proposed objective function thanks to the use of local shape constraints instead of a global shape description. Nevertheless, these methods did not get a lot of attention in the literature of medical image analysis. Instead, these local methods have been applied for different computer vision problems where a number of features in the image are recognized. For example Felzenszwalb used his method for the detection of persons and facial features from pictures [26]. Coughlan [25] applied his method for the segmentation of hands in pictures.

2.4 Related methods

This section describes the methods most related to the model-based segmentation approach developed in this thesis: active shape models (Section 2.4.1), active appearance models (Section 2.4.2) and the pictorial structure algorithm put forward by Felzenszwalb and Huttenlocher [68] (Section 2.4.3).

2.4.1 Active shape models

The general ASM scheme consists of three components: a global shape model, a local gray-level appearance model and a search algorithm to match the model to a new image. The original scheme of Cootes and Taylor described in [54] and the ASMOF algorithm with an improved intensity model (van Ginneken *et al.* [8]) are described briefly.

Shape Model

The shape of the object is described by a fixed number of labeled landmark points

$$\mathbf{l}_i = (x_i, y_i, z_i) \quad (2.1)$$

A training set is constructed by determining the landmark points in a set of s training images. A PDM is constructed by applying PCA to the training shapes

$$\mathbf{l} = (x_1, y_1, z_1, \dots, x_n, y_n, z_n). \quad (2.2)$$

First, the mean and covariance of the example shapes $\mathbf{l}^{(i)}$, $i = 1, \dots, s$ are computed:

$$\boldsymbol{\mu}_1 = \frac{1}{s} \sum_{i=1}^s \mathbf{l}^{(i)}, \quad (2.3)$$

$$\boldsymbol{\Sigma}_1 = \frac{1}{s-1} \sum_{i=1}^s (\mathbf{l}^{(i)} - \boldsymbol{\mu}_1)(\mathbf{l}^{(i)} - \boldsymbol{\mu}_1)^T. \quad (2.4)$$

Secondly, the eigenvectors and eigenvalues of $\boldsymbol{\Sigma}_1$ are computed. Retaining the eigenvectors ϕ_i corresponding to the t largest eigenvalues λ_i enables the approximation of a shape \mathbf{l} that belongs to the class of shapes represented by the training set:

$$\mathbf{l} \approx \boldsymbol{\mu}_1 + \boldsymbol{\Phi} \mathbf{b} \quad (2.5)$$

with $\boldsymbol{\Phi} = [\phi_1 \dots \phi_t]$. The number of principal components t is determined by f_v , the amount of variation in the training shapes to include in the model. The vector \mathbf{b} represents the model parameters and is computed as

$$\mathbf{b} = \boldsymbol{\Phi}^T (\mathbf{l} - \boldsymbol{\mu}_1) \quad (2.6)$$

This way, the sum of squared differences (SSD) between \mathbf{l} and the reconstructed shape (Eq. 2.5) is minimized. The values of \mathbf{b} are restricted to lie in a range $\pm a\sqrt{\lambda_i}$ to generate only plausible shapes. The value of a usually lies between two and three.

Prior to PCA, the shapes can be aligned by translating, rotating and resizing them so as to eliminate irrelevant pose-related differences. This is usually obtained by minimizing the SSD between the corresponding landmark points using Procrustes analysis [70]. Omitting this global alignment requires a considerable higher value for f_v as the first modes of shape variation are usually associated with variations in size and position [8, 71].

Intensity model (Cootes and Taylor)

To augment the model with intensity information, knowledge about the local gray-level appearance around each landmark is captured. Two approaches are described.

The ASM method described in [54] builds intensity profiles as follows. First, the image intensities are sampled at k pixels on either side of the shape contour at which the landmark is located with a fixed step size perpendicular to the contour. To reduce the effect of global intensity offsets between images, the first derivative of the intensities is computed using finite differences. The resulting profiles are normalized such that the sum of absolute values is equal to one. Extracting the profile in each training image, results in a training set $\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(s)}$ describing the local gray-level appearance around the landmark of interest.

A statistical model of the local gray-level appearance is obtained by computing the mean $\boldsymbol{\mu}_f$ and covariance $\boldsymbol{\Sigma}_f$. The goodness of fit of a new profile \mathbf{f} ,

encountered when searching for a shape in an unseen image, is measured as the Mahalanobis distance between this profile and the mean profile:

$$d(\mathbf{f}) = (\mathbf{f} - \boldsymbol{\mu}_{\mathbf{f}})^T \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} (\mathbf{f} - \boldsymbol{\mu}_{\mathbf{f}}). \quad (2.7)$$

Intensity model (van Ginneken)

An improved non-linear appearance model was introduced by van Ginneken *et al.* [8]. For each landmark point, a nonlinear kNN-classifier is constructed to predict whether a point lies inside or outside the object.

The local image neighborhood of a pixel is described by a feature vector using the concept of *locally orderless images* (LOI) [9]. First, the original image is fed into a filter bank of multiscale Gaussian derivatives. Secondly, from this set of filtered images, feature images are extracted by taking the first few moments of the local distribution of image intensities around each location. The local image appearance of a specific pixel is then described by taking the intensity values of the feature images at the pixel of interest.

This way, a feature vector is constructed and is used to classify a pixel as background or as object. The kNN-classifier is constructed as follows. From each training image and for each landmark a square grid of $N_{\text{grid}} \times N_{\text{grid}}$ points with N_{grid} an odd integer, is centered at the landmark point. For each point on the grid, the feature vector is computed and labeled with 0 or 1 if the pixel lies outside or inside the object respectively. The number of features is limited to a predefined maximum with a sequential feature forward and backward selection algorithm yielding an optimal set of features.

The goodness-of-fit of a candidate location is then obtained by first defining a profile with length $2k + 1$ perpendicular to the contour centered at the candidate location (akin to the original ASM scheme). Secondly, for each point on the profile ($i = -k, \dots, 0, \dots, k$), the corresponding feature vector is fed into the classifier to determine the probability f_i that the corresponding location lies inside the object. The goodness-of-fit measure is then computed as the sum of absolute differences between the expected probability (0 or 1 for points outside or inside the object, respectively) and the predicted probability f_i for each point along the contour:

$$d(\mathbf{f}) = \sum_{i=-k}^{-1} f_i + \sum_{i=0}^k (1 - f_i) \quad (2.8)$$

Several profile step sizes, image scales and grid spacings are used to build local intensity models for multiple resolutions.

Search Method

The shape and intensity models are applied to a new image in an iterative coarse-to-fine optimization algorithm initialized with the mean shape. Each iteration consists of two steps: (1) updating the landmark points using the local appearance models and (2) fitting the shape model to the updated target

points.

In the first step, each landmark is moved along a line perpendicular to the contour to n_s positions at either side with a step size in accordance to the current resolution level. Each of the $2n_s+1$ locations is evaluated by computing the goodness-of-fit of the profile centered at that location and perpendicular to the current contour.

After each landmark is moved to the best fitting point, the shape model is fitted to the updated points. This is repeated for a fixed number of iterations or till some convergence criterium is reached. This procedure is repeated for each resolution.

2.4.2 Active appearance models

The input for an AAM [56] is the same as for an ASM: a set of training images augmented with landmark points marked in each image. AAM builds a combined model of shape and appearance.

Combined shape and intensity model

Whereas ASM does a PCA analysis of only the shape \mathbf{l} , an AAM builds a combined model of both shape \mathbf{l} and texture \mathbf{f} . The texture \mathbf{f} holds the intensities of the image after the image was made shape-free. This is obtained by warping the image such that the corresponding shape \mathbf{l} is transformed onto the mean shape $\mu_{\mathbf{l}}$.

The appearance model has parameters \mathbf{c} and generates a shape instance \mathbf{l} with corresponding texture \mathbf{f} as follows:

$$\mathbf{l} = \mu_{\mathbf{l}} + \Phi_{\mathbf{l}}\mathbf{c} \quad (2.9)$$

$$\mathbf{f} = \mu_{\mathbf{f}} + \Phi_{\mathbf{f}}\mathbf{c} \quad (2.10)$$

where $\mu_{\mathbf{l}}$ is the mean shape, $\mu_{\mathbf{f}}$ is the mean texture and $\Phi_{\mathbf{l}}$, $\Phi_{\mathbf{f}}$ follow from the eigen-analysis performed on the training set. If the generated texture \mathbf{f} (Eq. 2.10) is warped onto the generated geometry \mathbf{l} (Eq. 2.9) a new image instance is synthesized.

Search method

In order to segment an unseen image \mathbf{I} , the AAM search method tries to find parameters \mathbf{c} that minimize the difference between the new image and the synthesized image.

The error measure of a specific \mathbf{c} is computed as follows. Firstly, the model (Eq. 2.9 and Eq. 2.10) is used to suggest a shape \mathbf{l} and texture \mathbf{f}_m . Subsequently, the suggested shape is used to sample the image texture \mathbf{f}_i . The error is computed as

$$E(\mathbf{I}, \mathbf{c}) = \|\mathbf{f}_m(\mathbf{c}) - \mathbf{f}_i(\mathbf{I}, \mathbf{c})\|^2 \quad (2.11)$$

Hence, the AAM search is driven by differences in modeled texture and observed texture.

It is known that the AAM search problem is not easily solved due to its non-convex nature. Cootes and Taylor [56] proposed to learn how the problem is solved during the training phase. For a specific training image, there exist model parameters \mathbf{c} that predict the texture in the image \mathbf{f} . They observed that an error in the difference vector $\delta\mathbf{f}$ is related to an error in the model parameters $\delta\mathbf{c}$. The assumed linear relationship

$$\delta\mathbf{c} = \mathbf{A}\delta\mathbf{f} \quad (2.12)$$

is learnt during the training phase using linear regression.

Optimization is done iteratively: a current \mathbf{c} enables to compute the texture difference $\delta\mathbf{f}$. The model parameters are then updated $\mathbf{c} \rightarrow \mathbf{c} - \delta\mathbf{c}$ with $\delta\mathbf{c}$ computed from Eq. 2.12. This is repeated till no improvement is made to the error (Eq. 2.11). Usually, this iterative scheme is run for multiple resolution levels.

AAM extensions

The original AAM scheme has some problems when applied for medical image segmentation tasks. First, the quadratic error measure (Eq. 2.11) makes AAM sensitive to outliers. In this context an outlier means that the image contains intensity patterns that are not explained by the model. For example a local disturbance caused by a pathology entangles the search for a correct solution. To cope with this, alternative error measurements were proposed [58, 72] to make the AAM search more robust. Beichel *et al.* [60] proposed an alternative search method to increase robustness.

Another problem encountered when applying the standard AAM scheme for medical applications was mentioned in the literature [11, 58]. As the standard AAM scheme only considers the object's interior for constructing the texture vectors, and because the inside of the anatomical objects often does not contain valuable and discriminating information, the search is likely to end up at the inside of the object. Stegmann *et al.* [58] proposed to include intensity information that lies at the object's outside. van Ginneken *et al.* [11] added texture information by adding intensities sampled along the contour normals at each landmark point. They proved that this approach resulted in a significantly better segmentation performance for the delineation of anatomical structures in chest radiographs.

2.4.3 Pictorial structures

The work of Felzenszwalb and Huttonlocker on object recognition using pictorial structures [26, 73] is summarized here.

Model

The object is represented as a set of nodes or landmark points \mathbf{I}_i and a set of connections $\{i, j\}$ between the nodes in a tree configuration (Fig. 2.4). In order

to find an instance $\mathbf{l} = \{\mathbf{l}_1, \dots, \mathbf{l}_n\}$ of the object in an image \mathbf{I} , two costs are defined. First, the match cost $d_i(\mathbf{I}, \mathbf{l}_i)$ measures how well a point \mathbf{l}_i matches to the image data. Secondly, a pairwise shape cost $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$, defined for each connection $\{i, j\} \in \mathcal{E}$, measures the plausibility of the relative geometrical configuration of \mathbf{l}_i with respect to \mathbf{l}_j . The definition of the match cost and geometrical cost leads to the following energy function

$$E(\mathbf{I}, \mathbf{l}) = \sum_{i=1}^n d_i(\mathbf{I}, \mathbf{l}_i) + \sum_{\{i,j\}} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \quad (2.13)$$

Search method

A proper segmentation \mathbf{l}^* is found as the shape instance that minimizes Eq. 2.13:

$$\mathbf{l}^* = \arg \min_{\mathbf{l}} E(\mathbf{I}, \mathbf{l}) \quad (2.14)$$

As each landmark point \mathbf{l}_i can be placed at m locations with m the number of pixels in the image, Eq. 2.14 is a discrete optimization problem. Dynamic programming [10] is used to find a global optimum of this cost function.

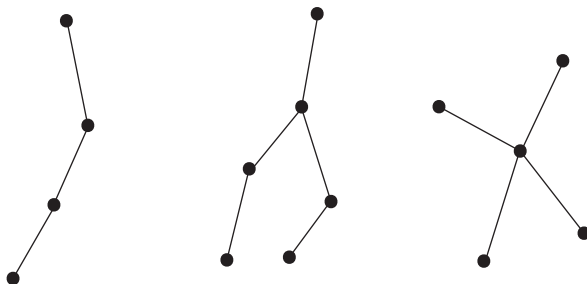


Figure 2.4: Some example graphs in a tree-like configuration.

2.5 Motivation

Here the pillars of the proposed segmentation scheme are explained and motivated.

2.5.1 Landmarks

As previously mentioned, there is no standard method for image segmentation as different applications need different strategies. Obviously, if the tissue types of the anatomical objects of interest have distinguishing photometry features, low-level segmentation and rule-based algorithms might be considered. However, due to the complex nature of the images and objects, most medical applications require the use of a model.

The type and the amount of prior knowledge that a model can include, largely depends on how the object is described. The shape-based approaches, certainly the ones with landmarks, have a higher modeling capacity than the intensity based methods. However, this does not mean that a shape-based algorithm should be preferred by all means. One of the conditions when using landmarks, is that a one-to-one mapping must exist between all shape instances of the same anatomical structure. Vessel trees for example do not satisfy this condition as the number of branches and bifurcations in the tree is not known in advance. Another example is brain tissue classification (e.g, [42]) from MR datasets. For these applications, pixel classification methods should be considered instead.

Nevertheless, if the correspondence condition is satisfied, which is true for many anatomical objects, segmentation with landmark-based models is an appropriate choice because of a number of reasons. Firstly, a landmark-based shape representation, allows to learn all the available shape characteristics from a set of examples. Moreover, such an algorithm is highly generic as it can be adapted easily to a new application by replacing the examples in the training set. Another advantage of the landmark approaches is that the image appearance can be modeled both locally, for each landmark, or globally. Hence, a uniform appearance distribution along the object boundary is not required. As a result, among the shape-based approaches, segmentation using ASMs and AAMs became the most popular strategy in the medical domain.

2.5.2 Problems with global PDMs

However, during the years, some shortcomings related to global PDMs were addressed in the literature (e.g.,[5, 6]).

Local instead of global optimization

The first drawback is rather a problem of the matching algorithms than it is a limitation inherent to the PDM itself. ASM is the best known and most widely used technique where a global PDM is matched to an image. Unfortunately, the ASM search is prone to find a solution which is only locally optimal. As in each iteration target landmark points are searched for in a local neighborhood of the current estimate of each landmark location, a sufficiently accurate initialization needs to be provided for the scheme to converge to the correct shape. This is illustrated in Fig. 2.5 for lung field segmentation in a chest radiograph.

Also the fact that the ASM optimization algorithm applies the intensity model and the shape model alternatingly instead of simultaneously, might cause the algorithm to get stuck in a local optimum. In each iteration, the landmark locations are updated using the intensity model only and subsequently a new ASM shape is fitted through the target points. If the intensity model selects the wrong location for a landmark, the shape model is misled. An example of this is shown in Fig. 2.6.

As a result, a number of ASM variants were suggested to avoid the search to get stuck into a local minimum. A more robust intensity model was proposed by van Assen *et al.* [22] by using image patches instead of pixel profiles. Behiels *et al.* [20] and Tianly *et al.* [74] proposed a smoothness constraint imposed on the displacement of neighboring landmark points to avoid the intensity model to select a wrong candidate. This makes the ASM approach more robust to local optima, although extra improvement could be made if the shape and intensity model would be applied simultaneously. The problem of the alternating use of the two models was also mentioned by Gleason *et al.* [75]. They proposed a probabilistic based segmentation scheme that iteratively optimizes a maximum a posteriori (MAP) objective function taking both shape and intensity knowledge into account [6, 65]. Behiels *et al.* [20] introduced a similar MAP algorithm although it seemed less effective than the ASM with smoothness constraint. In spite of all these improvements, the global shape model keeps complicating the search for the global optimum.

Flexibility

As the global PDM is built from examples in a training set, only shape instances can be reproduced that have the same variational modes as the example shapes. Consequently, if the training set is too small, several new shapes can not be reproduced properly. If such a shape is segmented, and even if the intensity model attracts the shape to the correct image boundaries, the shape model will pull the shape away from the correct solution. To avoid this, a sufficiently large representative training set is needed. However, this is for several applications not acceptable as it takes a large effort to obtain the examples, certainly for 3D applications.

To overcome this lack of flexibility, some researchers (e.g., [5, 76, 77]) proposed to superimpose local degrees of freedom onto the global shape description. They combined a global model with a constrained free-form deformable model to gain extra flexibility. This allows the shape template to fit more accurately to unseen images, even when the shape has variational modes that do not occur in the training set.

2.5.3 Local graphical templates

To cope with the limitations of the global PDMs, instead of superimposing local degrees of freedom onto a global shape description, alternatively the graphical deformable templates [23–26, 68] (Section 2.3.2) could be used. As these methods only consider local shape constraints they are much more flexible. Moreover, some of them guarantee a global optimum of the posed optimization problem.

Nevertheless, these methods are rarely used in the medical domain. As no extensive validation study exists for medical applications, these algorithms and their performance is rather unknown to the medical community. A first possible reason is that these algorithms are often not directly applicable to medical

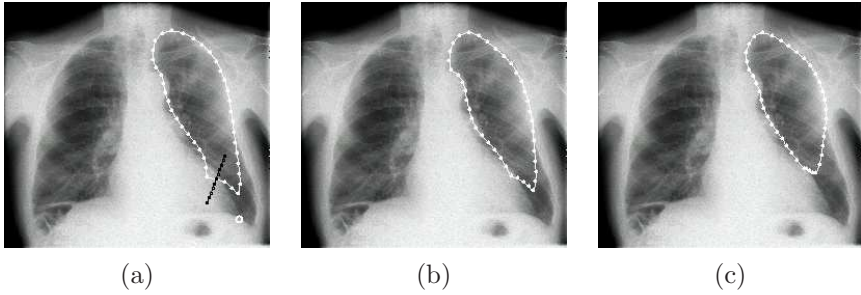


Figure 2.5: Illustration of an ASM shortcoming with a lung field segmentation example. The large distance between the true location of the landmark point indicated with a white circle (a) and the search line perpendicular to the initial contour (mean shape) avoids the shape to converge to the true solution. The contour after one iteration and the final contour are shown in (b) and (c) respectively.

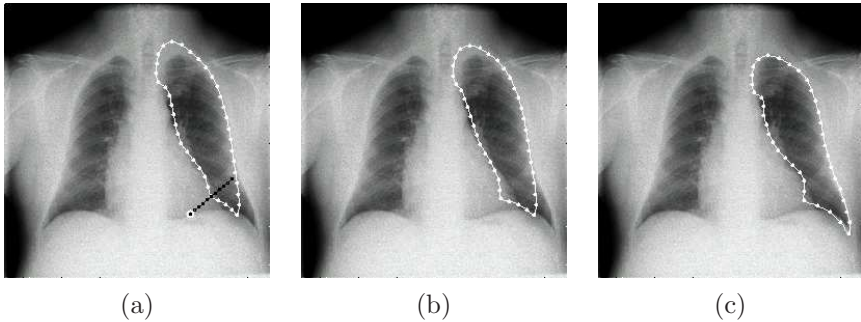


Figure 2.6: The gray-level model searches for a better landmark location by computing the goodness of fit (Eq. 2.7) of 11 ($n_s = 5$) candidate locations as shown in (a). The point that best fits the gray-level model, indicated with a white circle, is a bad choice from a shape point of view and distracts the shape from the true solution. The contours after the current iteration and the final solution are shown in (b) and (c) respectively.

applications. For example the pictorial structures algorithm described in Section 2.4.3 uses tree-like object representations and is therefore not capable of segmenting surfaces from volumetric images. In computer vision, these methods are considered as a solution for object recognition problems where features in pictures are detected, rather than as a method for segmentation. A second possible explanation is that many researchers believe that a global shape model is a requirement towards robust segmentation of anatomical structures. With good reason one can indeed wonder whether using only local statistical knowledge suffices.

In this thesis a segmentation scheme will be presented that falls in the category of the deformable graphical templates. A model is proposed that learns local

shape and appearance characteristics from examples and is generally applicable for the segmentation of anatomical objects from both 2D and 3D images.

2.6 Conclusion

Choosing an appropriate algorithm is a first and crucial step towards image segmentation. In this chapter a short overview is given of the best known segmentation algorithms together with their application fields. These methods are divided according to their use of prior knowledge: low-level methods versus model-based segmentation schemes.

Anatomical objects often have a large shape variability. Moreover, their appearance in medical images is often variable and noisy. Consequently, a model that maximally captures prior knowledge is required for a robust and accurate segmentation.

The amount and the nature of the prior knowledge that can be modeled largely depends on how the object is represented. A strong approach is to use landmark points because of its large modeling capacity. As a result, active shape and active appearance models are very popular in medical image analysis. However, there are a few limitations. Their use of a global shape model complicates the search for an optimum, and moreover a large number of training examples is needed to make the model flexible enough towards a sufficiently accurate segmentation.

The deformable graph methods capture local shape characteristics and are thereby much more flexible. Nevertheless, they haven't received much of attention for medical segmentation tasks. In this thesis, such a deformable graph model will be presented and its validity for the segmentation of anatomical objects, both in two as in three dimensions, will be verified.

Chapter 3

Modeling image objects

3.1 Introduction

The automated segmentation of anatomical structures requires prior knowledge about the geometry of the objects and knowledge related to how the objects appear in the image. A mathematical description of prior geometry knowledge is referred as a shape model. A description of how the image data typically looks like in that part of the image where the object is located, is called an intensity model. Also the expressions gray-level appearance model or even shorter appearance model are commonly used.

Different strategies exist for constructing the models. One possibility is to use a mathematical model that simulates the physical and bio-mechanical properties of the anatomical object of interest. Also a set of rules (Section 2.2.3) can be developed to incorporate the available prior knowledge. Unfortunately, these approaches have the limitation to be rather dedicated to one single application.

A much more generic strategy, the one followed here, extracts the prior knowledge from examples. This learning or *training* process mimics the process that a human experiences when he learns to recognize objects in images. Before he is able to segment a specific anatomical structure, he first needs to learn how the object typically appears by observing a set of annotated example images. Automated algorithms that learn their model from training data are categorized as *supervised* methods.

Many of these supervised methods applied for medical purposes perform a principal component analysis of the global shape vector to learn the shape characteristics. This is a strong and natural approach as it learns all the existing correlations that can be observed in a set of example shapes. However, such a global shape description has the consequence that if the location of only one point on the shape is changed, every other point feels a direct influence. This is somewhat unnatural. Consider the example of a liver tumor. It is acceptable to assume that the tumor will change the liver geometry only in its neighborhood and that it will not interact with the shape of the liver at

the opposite side. To overcome the lack of flexibility of such a global shape description, a localized model based on Markov random field theory will be proposed that only considers local shape dependencies.

To understand how the chapter is organized, a little explanation is needed about the statistical nature of the model and how to apply it for the segmentation of an unseen image. An image \mathbf{I} and the corresponding segmentation \mathbf{l} are considered as correlated random processes. The segmentation task can be interpreted as finding the shape \mathbf{l} in the image \mathbf{I} that maximizes the posterior probability $p(\mathbf{l}|\mathbf{I})$. Applying Bayes' rule

$$p(\mathbf{l}|\mathbf{I}) = \frac{p(\mathbf{l})p(\mathbf{I}|\mathbf{l})}{p(\mathbf{I})} \quad (3.1)$$

separates the process into a shape prior $p(\mathbf{l})$ that reflects the plausibility of the shape and an image appearance model $p(\mathbf{I}|\mathbf{l})$ that expresses how well the shape fits to the given image. The probability $p(\mathbf{I})$ does not need to be estimated as we are only interested in how $p(\mathbf{l}|\mathbf{I})$ varies with \mathbf{l} . How this optimization problem is solved will be explained in Chapter 4.

This chapter explains how the prior knowledge, embodied by the distributions in Eq. 3.1, is learnt from examples. The chapter is organized as follows (Fig. 3.1). Prior to estimating the shape and appearance distributions, a mathematical representation of the shape \mathbf{l} (Section 3.2) and a set of training examples (Section 3.3) are needed. Section 3.4 explains how probabilities are converted into energies. The shape prior $p(\mathbf{l})$ is explained in Section 3.5 and the intensity distribution $p(\mathbf{I}|\mathbf{l})$ is revealed in Section 3.6. Finally, these models are merged into one global model that combines shape and intensity characteristics (Section 3.7).

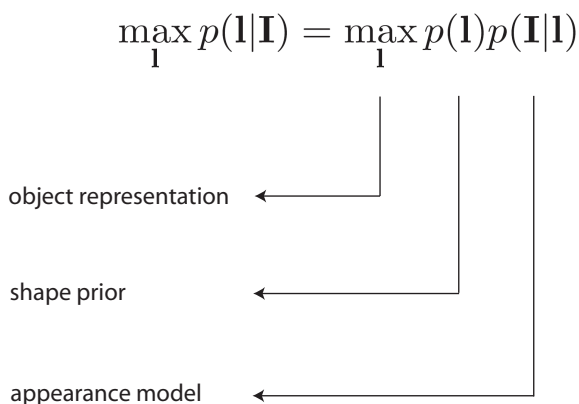


Figure 3.1: Components for model construction: an object representation, a shape prior and an appearance model.

3.2 Object representation

Landmarks and edges

The object is represented as a graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}) \quad (3.2)$$

- where the vertices $\mathcal{V} = \{1, \dots, n\}$ refer to landmark points $\mathbf{l}_1, \dots, \mathbf{l}_n$ defined on fixed positions on the shape,
- and the edges \mathcal{E} contain t unordered pairs of distinct vertices $\{i, j\}$ that express landmark interactions.

Consider the examples in Fig. 3.2 and 3.3. The landmark points can be placed at the object boundaries resulting in contours (Fig. 3.2(a,f)) and surfaces (liver in Fig. 3.2(b)) for respectively two and three dimensional objects. Alternatively, the landmark points can be defined at the inside of the object (lung in Fig. 3.2(c)) or landmarks can be used to annotate characteristic features in the image as in Fig. 3.2(d,e).

The exact meaning of the edges will become clear when the shape prior $p(\mathbf{I})$ is introduced (Section 3.5), but it is already useful to reveal what they are meant for. Edges will be defined between those landmarks of which the locations have a direct statistical interaction. When landmarks describe a 2D curve along an object boundary, it is straightforward to define edges between successive points (Fig. 3.2(a,f)). Extending this representation to three dimensional shapes results in surface meshes (Fig. 3.3) where edges are defined between direct neighbors along the shape border. Edges do not necessarily coincide with object borders. For example, the graph in Fig. 3.2(d) assumes that there is an interaction between neighboring bones of the same finger. The graph in Fig. 3.2(e) also takes dependencies between different fingers into account.

Neighborhood system

Instead of defining a set of edges \mathcal{E} , the interrelationship between the landmarks can be formulated in terms of a neighborhood system \mathcal{N}

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{V}\} \quad (3.3)$$

where \mathcal{N}_i is the set of sites neighboring i . The neighborhood system satisfies the following requirements:

1. a vertex is not neighboring to itself: $i \notin \mathcal{N}_i$
2. the neighboring relationship is mutual: $i \in \mathcal{N}_j \Leftrightarrow j \in \mathcal{N}_i$

There is a one-to-one mapping between \mathcal{E} and \mathcal{N} :

$$j \in \mathcal{N}_i \Leftrightarrow \{i, j\} \in \mathcal{E} \quad (3.4)$$

Both the edges \mathcal{E} as the neighborhood system \mathcal{N} will be used in this text.

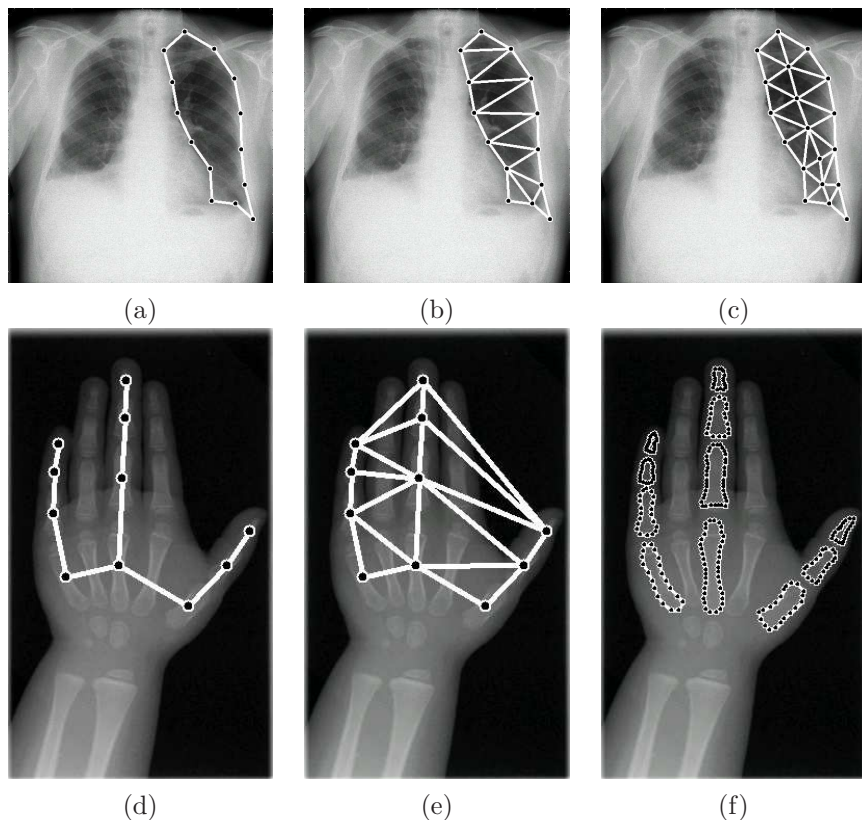


Figure 3.2: The shape of anatomical objects represented as graphs. The shapes can be described with landmark points defined along the object boundary (a,b,f). Landmark points can also be located at specific points inside the objects (c) or at specific features in the image (d,e).

Cliques

Another term commonly used in graph theory needs to be explained. A *clique* c is a subset of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ such that every two vertices in the clique are neighbors. The maximum clique size of a graph will have an important meaning when the shape distribution is discussed.

The maximum clique size of a contour (Fig. 3.2(a,f)) or a tree-like configuration (Fig. 3.2(d)) is two. A triangulated graph has cliques up to order three (Fig. 3.2(b,c,e)).

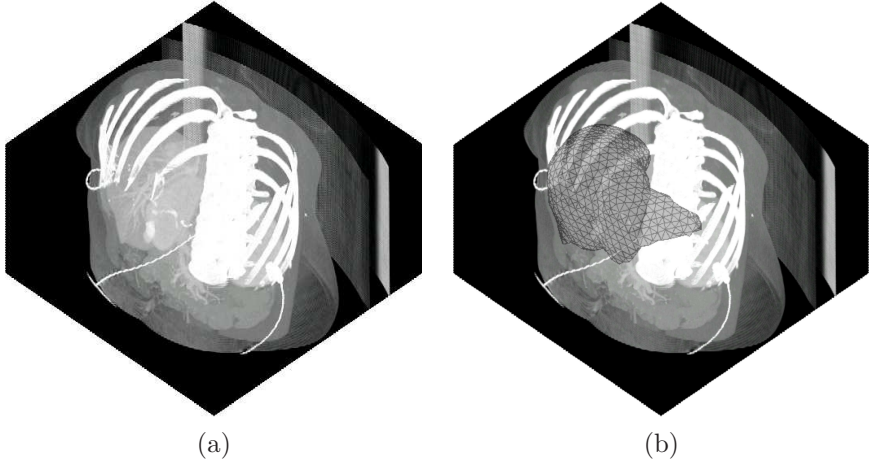


Figure 3.3: The graph representation is appropriate for modeling 3D objects. The liver clearly visible in the contrast enhanced CT image (a) can be described as a triangulated surface mesh (b).

3.3 Training set

The training set consists of s example images with corresponding ground-truth shapes:

$$\{(\mathbf{I}^{(k)}, \mathbf{l}^{(k)})\}_{k=1}^s \quad (3.5)$$

Each shape $\mathbf{l}^{(k)} = \{\mathbf{l}_1^{(k)}, \dots, \mathbf{l}_n^{(k)}\}$ consists of n landmark points. Setting up such a set yields different stages.

3.3.1 Ground-truth segmentations

Firstly, a set of ground-truth segmentations is obtained by segmenting a set of images using manual or semi-automated techniques. In this thesis, for some applications we make use of publicly available data. For example the SCR database [11, 12] contains 247 posterior-anterior chest radiographs together with manual delineations of lung fields, heart and clavicles. A free database that supplies training material for 3D liver segmentation from computed tomography datasets is available at [78].

Whether the ground-truth data is obtained from such a public database or obtained with manual segmentation tools, in general the shapes are not yet represented as a set of landmark points. A manually delineated contour consists of an arbitrarily number of points clicked along the object boundary rather than the required landmark points. Other semi-automated segmentation tools fill regions in the image to define the object's location and produce labeled images. Again, no information about corresponding points is available yet. A training set is only complete if the point correspondence problem is solved.

3.3.2 Correspondences

Finding corresponding points on 2D objects is often done by hand. To do this efficiently, a number of specific anatomically characteristic points is annotated on each shape instance first. Additionally, a number of pseudo-landmarks are defined by equidistantly sampling points along the object boundary between the characteristic landmarks.

Unfortunately, this manual technique is for most 3D applications not acceptable as it would become too cumbersome and would take too much time. As a result, automated algorithms that search corresponding points on instances of anatomical structures have become very popular and well studied in the domain of medical image analysis. A complete survey of the existing approaches to search the correspondences falls out of the scope of this work. Instead, only a limited number of important and relevant methods are reported here.

Many published methods use a point registration algorithm to deform a template shape onto a target shape. The warped points of the template shape are then considered as the landmark points of the target shape.

A famous algorithm in this category is the iterative closest point (ICP) algorithm proposed by Besl *et al.* [79]. For each point on the target shape, the closest point to the template surface is searched. Subsequently, a transformation is computed that brings the target points closer to the corresponding closest points of the template shape. This procedure is repeated until the change in the objective function falls below a preset threshold.

In Chapter 6, a similar method [80] will be used to warp a template surface onto a target surface. First, the target shape is converted into a distance map. This *distance* image contains the closest distance to the surface at each pixel position. The template shape is then iteratively transformed by letting each landmark point evolve with a small step in the direction of the gradient of the distance map. This way, each landmark point of the template is moved closer to the target surface. This approach is computationally less expensive compared to the ICP method as the closest points do not have to be computed explicitly in every iteration.

The minimal description length (MDL) [81] approach to the point correspondence problem, introduced by Davies *et al.* became recently very popular and therefore needs to be mentioned. They proposed to compute the quality of a set of correspondences using an information theoretic measure. The idea is that the amount of information needed to describe the shape statistics extracted from a set of correspondences should be as small as possible. The compactness of the shape model is computed as the description length that is needed to encode the model.

A comparative study [82] showed that MDL outperforms several other strategies. At the other hand, one should be aware of the computationally expensive nature of the method. Recently, Heimann *et al.* [83] presented an optimized procedure for minimizing the MDL cost function that leads to a more efficient search and accordingly needs a considerably smaller computation time.

3.3.3 Global alignment

For reasons that will become clear in Section 3.5.3, the training shapes and corresponding images need to have similar orientations prior to model construction and model fitting (Section 3.5.3). In medical images, this requirement is often satisfied automatically thanks to the acquisition protocol. This is for example the case for radiographic images of the chest (Fig. 3.4).

If large pose differences of the shapes in the training images are present, a coarse affine alignment of the images is needed to correct for differing orientations and scales.

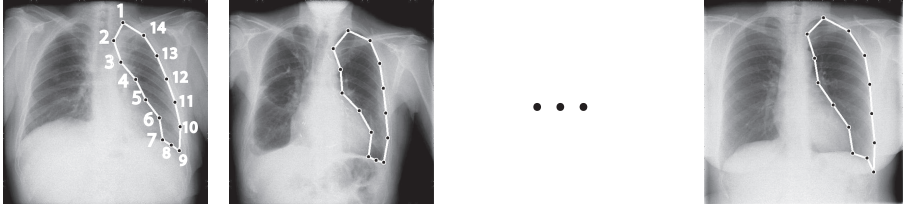


Figure 3.4: The lung fields in the radiographic images have similar orientation and scale due to the acquisition protocol.

3.4 Energy space

In this thesis several random variables, random vectors and random fields (Section 3.5.1) will be encountered. Instead of describing them with their probability density functions (PDFs), they will be described in the log-space as follows. Consider a random variable, random vector or random field $\mathbf{x} \in \mathbb{X}$ and the corresponding PDF $p(\mathbf{x})$. If one assumes that $p(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathbb{X}$ then one can write

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{1}{T}E(\mathbf{x})\right) \quad (3.6)$$

with T a strictly positive scalar and Z a normalizing constant

$$Z = \sum_{\mathbf{x} \in \mathbb{X}} \exp\left(-\frac{1}{T}E(\mathbf{x})\right) \quad (3.7)$$

We will refer to $E(\mathbf{x})$ as the energy of \mathbf{x} . If T is chosen, the energy $E(\mathbf{x})$ is determined except for a constant. Changing this constant alters the value of Z . A likely sample \mathbf{x} (high probability) corresponds to a low energy and vice versa: a sample with high energy is unlikely to occur (Fig. 3.5).

3.5 Shape model

A localized point distribution model (PDM) is proposed. First, a short introduction (Section 3.5.1) to Markov random field (MRF) theory [7] is given, as

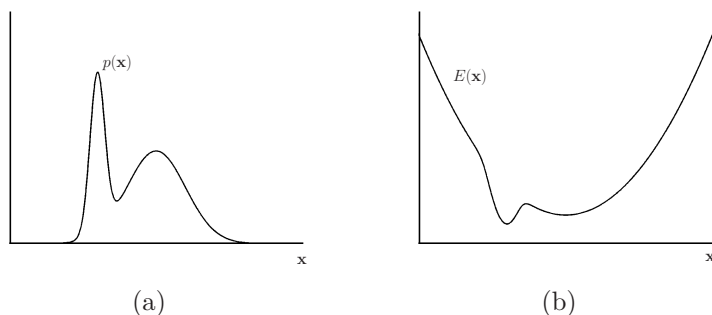


Figure 3.5: The PDF (a) of a random variable can be converted to the energy space (b). A high probability corresponds to a low energy and vice versa: a sample with a high energy is unlikely to occur.

this is the basis on which the PDM is built. A shape prior and corresponding energy function will be derived in Sections 3.5.2 and 3.5.3. Section 3.5.4 explains how the parameters of the energy function are estimated from a set of training shapes.

3.5.1 Markov random fields

Let $\mathcal{V} = \{1, \dots, n\}$ be a set of sites and $X = \{X_1, \dots, X_n\}$ a set of random variables (or random vectors) defined on \mathcal{V} , in which each variable X_i takes a value x_i in \mathbb{X} . A joint event $\{X_1 = x_1, \dots, X_n = x_n\}$ will shortly be noted as $\mathbf{x} = \{x_1, \dots, x_n\}$ and the probability will be written as $P(\mathbf{x})$ for a discrete random variable or as $p(\mathbf{x})$ for a continuous variable.

Let \mathcal{N} be a neighborhood system defined on \mathcal{V} , then X is said to be a Markov random field on \mathcal{V} with respect to \mathcal{N} if and only if the following two conditions are satisfied:

$$p(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in \mathbb{X}^n \quad (\text{positivity}) \quad (3.8)$$

$$p(x_i | x_{\mathcal{V}/\{i\}}) = p(x_i | x_{\mathcal{N}_i}) \quad (\text{Markovianity}) \quad (3.9)$$

The Markovianity property expresses that a site only interacts with its neighbors.

Gibbs energy

The Hammersley-Clifford theorem [7] states that an MRF is equivalent with a Gibbs random field (GRF). A GRF is a random field X defined by a joint PDF written as a Gibbs distribution:

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{1}{T}E(\mathbf{x})\right) \quad (3.10)$$

where Z is a normalizing constant called the partition function and T is a constant called the temperature of the Gibbs distribution. The energy function

$$E(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) \quad (3.11)$$

is a sum of clique potentials $V_c(\mathbf{x})$ over all possible cliques \mathcal{C} . The value of $V_c(\mathbf{x})$ only depends on the sites in c . Thanks to the Hammersley-Clifford theorem, the joint PDF of an MRF can be expressed in terms of potential functions defined on the cliques. Formula 3.10 enables to work with energies instead of probability densities: a field with a low probability corresponds to a high energy and vice versa.

Markov conditional distribution

If the potential functions of the Gibbs distribution are known, the conditional probability can be computed as [7]:

$$p(x_i | x_{\mathcal{N}_i}) = \frac{\exp\left(-\frac{1}{T} \sum_{c \in \mathcal{A}} V_c(\mathbf{x})\right)}{\sum_{\mathbf{x}' \in \mathbb{X}} \exp\left(-\frac{1}{T} \sum_{c \in \mathcal{A}} V_c(\mathbf{x}')\right)} \quad (3.12)$$

where \mathcal{A} consists of the cliques containing i and where

$$\mathbf{x}' = \{x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n\} \quad (3.13)$$

equals \mathbf{x} at all sites except possibly i .

Shape as an MRF

The shape \mathbf{l} is modeled as an MRF on the vertices \mathcal{V} with respect to the neighborhood system \mathcal{N} :

$$p(\mathbf{l}_i | \mathbf{l}_{\mathcal{V} - \{i\}}) = p(\mathbf{l}_i | \mathcal{N}_i) \quad (3.14)$$

This implies that if the locations of a landmark's neighbors \mathcal{N}_i are known, then everything is known to predict the location of landmark \mathbf{l}_i . No extra information can be gained by knowing the location of other landmarks.

Consequently, if a landmark is moved to another position, only his direct neighbors will feel this change directly. However, this does not mean that two landmarks that are not neighbors are statistically independent. There is interaction between them, but only indirectly via other landmarks.

3.5.2 Shape energy

The Hammersley-Clifford theorem delivers the following formula for the probability density function of the shape:

$$p(\mathbf{l}) = \frac{1}{Z_1} \exp\left(-\frac{1}{2} E_1(\mathbf{l})\right) \quad (3.15)$$

with Z_1 a normalizing constant. Without loss of generality, the temperature is chosen to be $T = 2$.

The theorem states that the *shape energy* $E_1(\mathbf{l})$ is a sum of local energy functions over all possible cliques of the graph. For example, the shape energy of a contour can be expressed as a sum of potential functions that depend on couples of neighboring landmarks whereas the energy of a triangulated graph consists of local energies of maximum third order.

To find expressions for the potential functions, the PDF of trivial graphs consisting of one, two and three vertices will be analyzed.

First order

Consider a trivial shape that consists of only one landmark: $\mathcal{V} = \{1\}$ and $\mathcal{E} = \emptyset$. The proposed PDM aims at describing the shape variances rather than its relative location in space. Therefore, the distribution of a single landmark is considered to be invariant for translations. Hence, the PDF is uniform and also the energy is a constant. One is free to choose

$$E_1(\mathbf{l}_1) = 0 \quad (3.16)$$

Second order

Consider a graph that consists of two vertices and one edge: $\mathcal{V} = \{1, 2\}$ and $\mathcal{E} = \{\{1, 2\}\}$. The shape PDF can be written as follows:

$$p(\mathbf{l}_1, \mathbf{l}_2) = p(\mathbf{l}_1)p(\mathbf{l}_2|\mathbf{l}_1) \quad (3.17)$$

$$= p(\mathbf{l}_1)p(\mathbf{l}_2 - \mathbf{l}_1|\mathbf{l}_1) \quad (3.18)$$

Eq. 3.17 follows Bayes' rule. The conversion to Eq. 3.18 is trivial as the random occurrence of \mathbf{l}_2 given \mathbf{l}_1 is equivalent to the occurrence of $\mathbf{l}_2 - \mathbf{l}_1$ given \mathbf{l}_1 . As the shape PDF is modeled to be invariant for translations:

$$p(\mathbf{l} + \delta) = p(\mathbf{l}), \forall \delta \in \mathbb{R}^3, \quad (3.19)$$

it follows that the distribution of an edge vector $\mathbf{l}_2 - \mathbf{l}_1$ does not depend on its relative location in space:

$$p(\mathbf{l}_2 - \mathbf{l}_1|\mathbf{l}_1) = p(\mathbf{l}_2 - \mathbf{l}_1). \quad (3.20)$$

Hence, the shape PDF becomes

$$p(\mathbf{l}_1, \mathbf{l}_2) = p(\mathbf{l}_1)p(\mathbf{l}_2 - \mathbf{l}_1) \quad (3.21)$$

$$\propto p(\mathbf{l}_2 - \mathbf{l}_1) \quad (3.22)$$

Writing the distribution of an edge $\mathbf{e}_{ij} = \mathbf{l}_j - \mathbf{l}_i$ in the energy space

$$p(\mathbf{e}_{ij}) = \frac{1}{Z_{\mathbf{e}_{ij}}} \exp\left(-\frac{1}{2}d_{ij}(\mathbf{e}_{ij})\right) \quad (3.23)$$

enables the definition of the *edge energy* $d_{ij}(\mathbf{e}_{ij})$ that reflects the plausibility of the edge \mathbf{e}_{ij} . To keep the notations simple in the following sections, the energy will be written as $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$ instead of writing it explicitly as function of the edge vector.

An edge that significantly deviates from what is seen in a population will result in a high energy whereas an edge that is similar as in the population will yield a low-energy value. How these edge energies are learnt from examples will be explained in Section 3.5.4.

Finally, the shape energy of the second order graph becomes

$$E_1(\mathbf{l}_1, \mathbf{l}_2) = d_{12}(\mathbf{l}_1, \mathbf{l}_2) \quad (3.24)$$

Obviously, the energy function consists of one potential function of second order.

Third order

Assume a graph that consists of one triangle:

$$\mathcal{V} = \{1, 2, 3\} \quad (3.25)$$

$$\mathcal{E} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\} \quad (3.26)$$

Similar as for the second order potential, the joint PDF can be rewritten as

$$p(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = p(\mathbf{l}_1)p(\mathbf{l}_2|\mathbf{l}_1)p(\mathbf{l}_3|\mathbf{l}_1, \mathbf{l}_2) \quad (3.27)$$

$$= p(\mathbf{l}_1)p(\mathbf{l}_2 - \mathbf{l}_1|\mathbf{l}_1)p(\mathbf{l}_3 - \mathbf{l}_1|\mathbf{l}_1, \mathbf{l}_2 - \mathbf{l}_1) \quad (3.28)$$

$$= c^{\text{te}} \cdot p(\mathbf{e}_{12})p(\mathbf{e}_{13}|\mathbf{e}_{12}) \quad (3.29)$$

The conversion from Eq. 3.28 to Eq. 3.29 uses the translation-invariance assumption. Similar as for the edge distribution 3.23, an energy can be defined for the conditional likelihood:

$$p(\mathbf{e}_{13}|\mathbf{e}_{12}) \propto \exp\left(-\frac{1}{2}d_{13|12}(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)\right) \quad (3.30)$$

The *conditional edge energy* $d_{13|12}(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)$ reflects the likelihood of observing a certain \mathbf{l}_3 if both \mathbf{l}_1 and \mathbf{l}_2 are given.

The shape energy of a triangle becomes

$$E_1(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = d_{12}(\mathbf{l}_1, \mathbf{l}_2) + d_{13|12}(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) \quad (3.31)$$

Several combinations of Eq. 3.31 can be obtained by shifting the roles of \mathbf{l}_1 , \mathbf{l}_2 and \mathbf{l}_3 in Eq. 3.27. A better balanced formula is obtained by averaging all combinations:

$$E_1(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = \frac{1}{3} (d_{12} + d_{23} + d_{13} + d_{13|12} + d_{12|23} + d_{23|13}) \quad (3.32)$$

In accordance with the Hammersley-Clifford theorem, the triangle energy consists of potential functions up to third order. The same principle can be used to obtain the energy function of any graph configuration. For example, a graph that consists of only triangles (called a *triangulation*) can be written as a sum of potential functions up to third order.

3.5.3 Second order approximation

As follows from section 3.5.2, the complexity of the shape energy function is related to the maximum clique size of the graph. Unfortunately, incorporating these potentials in the energy function will have a large impact on the computational complexity of the model fitting. Therefore it is desirable that the energy can be written in terms of potential functions up to size two. An approximation is made here where third order (and higher) potentials are written as combinations of second order functions.

Consider again the triangle example and the following approximation:

$$p(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = p(\mathbf{l}_1)p(\mathbf{l}_2|\mathbf{l}_1)p(\mathbf{l}_3|\mathbf{l}_1, \mathbf{l}_2) \quad (3.33)$$

$$\approx p(\mathbf{l}_1)p(\mathbf{l}_2|\mathbf{l}_1)p(\mathbf{l}_3|\mathbf{l}_1) \quad (3.34)$$

This approximation ignores the influence of \mathbf{l}_2 on \mathbf{l}_3 , or in other words: it only takes the edges \mathbf{e}_{12} and \mathbf{e}_{13} into account. However, this approximation is only acceptable if the triangle instances have similar orientations and scales.

To understand this, assume that the orientation of the triangle would be random. The probability $p(\mathbf{l}_3|\mathbf{l}_1, \mathbf{l}_2)$ tells where \mathbf{l}_3 can be expected if both \mathbf{l}_1 and \mathbf{l}_2 are known. Knowing the direction of the edge $\mathbf{l}_2 - \mathbf{l}_1$ contains valuable information about the expected direction of the edge $\mathbf{l}_3 - \mathbf{l}_1$. Hence, the influence of knowing \mathbf{l}_2 can impossibly be ignored and thus the approximation $p(\mathbf{e}_{13}|\mathbf{e}_{12}) \approx p(\mathbf{e}_{13})$ does not make sense here. Similarly, if the shapes would have highly differing scales, the influence of a given edge on the distribution of another edge can not be ignored. Indeed, if the given edge vector has a large magnitude, one can expect that the other edge will have a large length too.

As to make the approximation acceptable, an alignment of the training shapes is required.

A better balanced approximation that takes all edges into account, is obtained by averaging the three variants of Eq. 3.34:

$$E_1(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = \frac{2}{3} (d_{12}(\mathbf{l}_1, \mathbf{l}_2) + d_{13}(\mathbf{l}_1, \mathbf{l}_3) + d_{23}(\mathbf{l}_2, \mathbf{l}_3)) \quad (3.35)$$

General shape energy

Assume a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with n vertices and t edges. The joint PDF can be approximated similar as in Eq. 3.34:

$$p(\mathbf{l}_1, \dots, \mathbf{l}_n) \approx p(\mathbf{l}_1)p(\mathbf{l}_2|\mathbf{l}_1)p(\mathbf{l}_3|\mathbf{l}_1, \mathbf{l}_2) \dots p(\mathbf{l}_n|\mathbf{l}_{n-1}) \quad (3.36)$$

with $i_j \in \mathcal{N}_{i_{j-1}}$, $j = 2, \dots, n$. The energy that corresponds to Eq. 3.36 consists of $n - 1$ second order potentials of the type $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$. An averaged expression that sums all the possible combinations of Eq. 3.36 consists of t energy terms $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$. If each edge has the same weight in the energy function, the global shape energy becomes

$$E_1(\mathbf{l}) = \frac{n-1}{t} \sum_{\{i,j\} \in \mathcal{E}} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \quad (3.37)$$

As the occurrence of the edge $\mathbf{l}_j - \mathbf{l}_i$ is equivalent with $\mathbf{l}_i - \mathbf{l}_j$, then it is also true that $d_{ij}(\mathbf{l}_i, \mathbf{l}_j) = d_{ji}(\mathbf{l}_j, \mathbf{l}_i)$. Hence, the energy Eq. 3.37 can be rewritten as

$$E_1(\mathbf{l}) = \frac{n-1}{2t} \sum_{i=1}^n \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \quad (3.38)$$

Fig. 3.6 applies this formula for a number of example graphs.

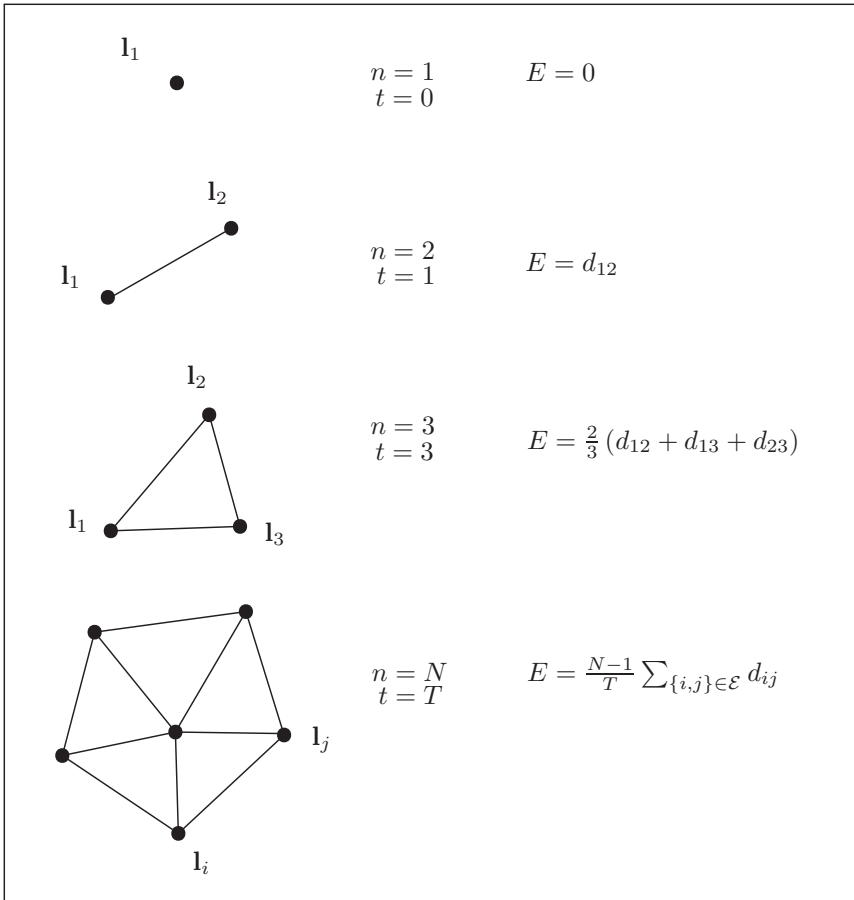


Figure 3.6: The shape energy of the graph can be approximated using only edge related energies d_{ij} .

Markov shape energy

The conditional distribution of the Markovianity property predicts were a landmark should lie given its neighbors and can be derived from the energy Eq. 3.38 using Eq. 3.12:

$$p(\mathbf{l}_i | \mathbf{l}_{\mathcal{N}_i}) = K(\mathbf{l}_{\mathcal{N}_i}) \exp \left(-\frac{n-1}{2t} \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \right) \quad (3.39)$$

with K a normalizing constant with respect to \mathbf{l}_i . The factor 2 arises here because $T = 2$. The conditional shape energy follows from Eq. 3.39:

$$E_{\mathbf{l}_i | \mathbf{l}_{\mathcal{N}_i}}(\mathbf{l}_i, \mathbf{l}_{\mathcal{N}_i}) = \frac{n-1}{t} \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \quad (3.40)$$

3.5.4 Learning from examples

In order to evaluate the proposed energy functions for a new shape instance, a probability density function (PDF) needs to be chosen that properly describes the edge distributions (Eq. 3.23 and Eq. 3.30) observed in the training set.

Edge energy

As the shapes have similar orientation and scale, an edge vector that connects two neighboring landmark points will have a more or less consistent length and direction. Hence, a multivariate gaussian distribution is a straightforward choice to describe the edge distribution (Eq. 3.23). It follows immediately that the defined edge energy $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$ is equal to the Mahalanobis distance [84] between the edge $\mathbf{l}_j - \mathbf{l}_i$ and the expected edge $\boldsymbol{\mu}_{\mathbf{e}_{ij}}$:

$$d_{ij}(\mathbf{l}_i, \mathbf{l}_j) = (\mathbf{l}_j - \mathbf{l}_i - \boldsymbol{\mu}_{\mathbf{e}_{ij}})^T \boldsymbol{\Sigma}_{\mathbf{e}_{ij}}^{-1} (\mathbf{l}_j - \mathbf{l}_i - \boldsymbol{\mu}_{\mathbf{e}_{ij}}) \quad (3.41)$$

with $\boldsymbol{\mu}_{\mathbf{e}_{ij}}$ and $\boldsymbol{\Sigma}_{\mathbf{e}_{ij}}$ respectively the mean and the covariance of the gaussian edge distribution.

The Mahalanobis distance measures how far a sample lies from its expectation value taking the modes of variation into account. For example, for a one-dimensional gaussian, the Mahalanobis distance is equal to the number of standard deviations that the sample deviates from its mean.

The unknown parameters in Eq. 3.41 are computed from the training shapes using the maximum likelihood (ML) criterion [85]:

$$\boldsymbol{\mu}_{\mathbf{e}_{ij}} = \frac{1}{s} \sum_{k=1}^s (\mathbf{l}_j^{(k)} - \mathbf{l}_i^{(k)}) \quad (3.42)$$

$$\boldsymbol{\Sigma}_{\mathbf{e}_{ij}} = \frac{1}{s-1} \sum_{k=1}^s (\mathbf{l}_j^{(k)} - \mathbf{l}_i^{(k)} - \boldsymbol{\mu}_{\mathbf{e}_{ij}})^T (\mathbf{l}_j^{(k)} - \mathbf{l}_i^{(k)} - \boldsymbol{\mu}_{\mathbf{e}_{ij}}) \quad (3.43)$$

The gaussian edge distributions estimated from the lung shapes in Fig. 3.4 are plotted in Fig. 3.7. The choice for a normal PDF seems convenient as it fits the training samples properly.

The Markov conditional distribution (Eq. 3.39) is illustrated for the hand shape. The vertices of the hand graph shown in Fig. 3.8 indicate the centers of 11 hand bones. The probabilities plotted in (b) and (c) show where the bone center is expected using only shape information. The red dots correspond to the true landmark locations.

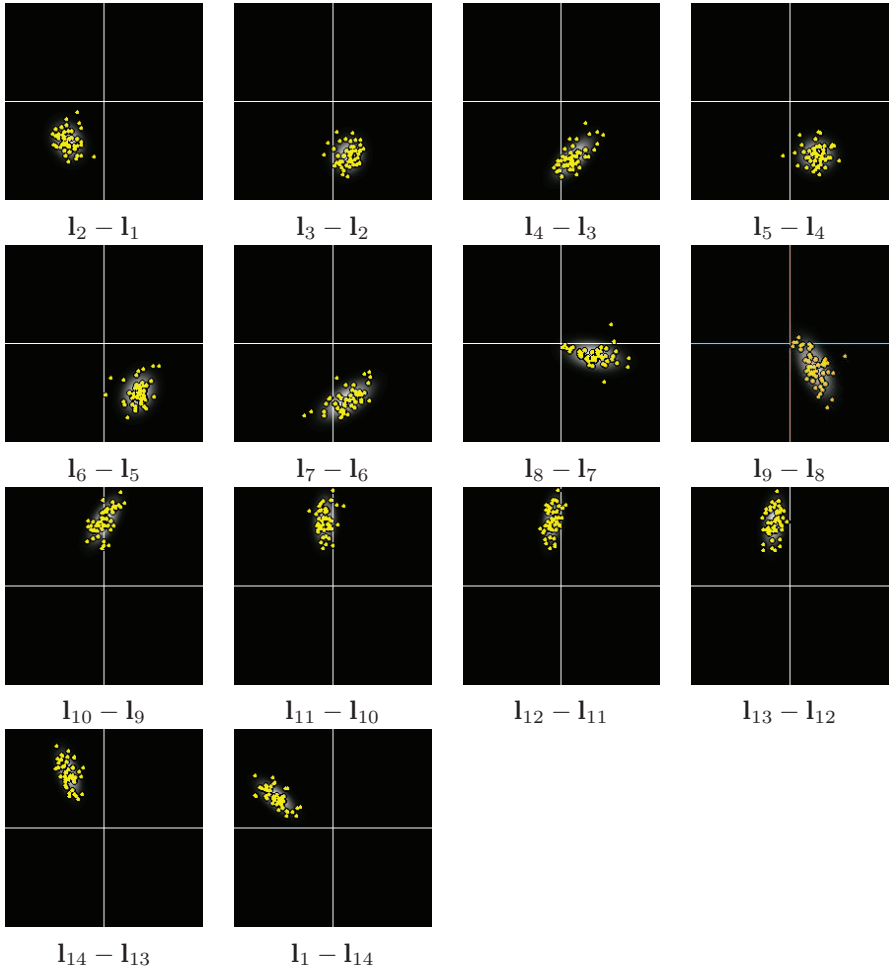


Figure 3.7: The estimated normal distributions (using Eq. 3.41) of the lung edges seem to describe the observed data properly. The Gaussian distributions are plotted as gray-levels where white corresponds to the maximum density of the distribution and black to a zero density. The training samples are plotted as yellow dots.

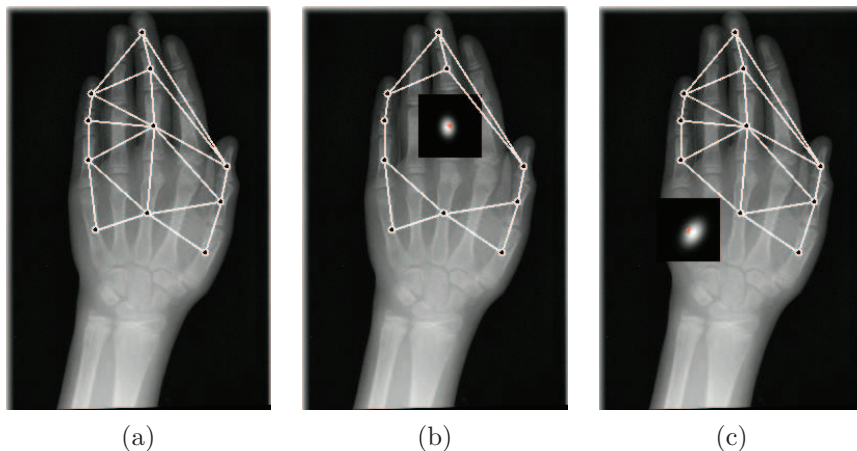


Figure 3.8: The conditional distribution of Eq. 3.39 is illustrated for the hand shape (a). The probabilities plotted in (b) and (c) show where the bone center is expected using shape information only. The red dots correspond to the true landmark locations.

Conditional edge energy

If the approximation presented in Section 3.5.3 is made, only second order potentials (Eq. 3.41) are needed. If at the other hand the exact energy is used and if the graph consists of cliques larger than order two, also the higher order potentials need to be estimated. In Chapter 5, a validation experiment will be performed where third order potentials are computed exactly. In that case, also the conditional edge distributions $p(\mathbf{e}_{ik}|\mathbf{e}_{ij})$ (Eq. 3.30) need to be estimated. The formulas can be found in Appendix A.

3.6 Intensity model

3.6.1 Model assumptions

The intensity distribution $p(\mathbf{I}|\mathbf{l})$ expresses the probability of observing the image \mathbf{I} given that the object of interest is located at \mathbf{l} . Two simplifications are made prior to estimating the model.

1. Firstly, the proposed model assumes that the shape has only an influence on the image photometry \mathbf{f} in the neighborhood of the shape \mathbf{l} . The image structure \mathbf{f} is obtained by extracting the local gray-level patterns \mathbf{f}_i around the landmarks \mathbf{l}_i using an intensity descriptor:

$$\mathbf{f}_i = F(\mathbf{I}, \mathbf{l}_i) \quad (3.44)$$

Hence, for each landmark a feature vector is extracted from the image that embodies how the image appears around that landmark point.

2. A second model assumption states that the landmark-individual patterns \mathbf{f}_i are statistically independent amongst one another. Consequently, one can write

$$p(\mathbf{I}|\mathbf{l}) \propto \prod_{i=1}^n p(\mathbf{f}_i) \quad (3.45)$$

with $p(\mathbf{f}_i)$ the probability that at the location of landmark i the local image pattern \mathbf{f}_i is observed. Note that the assumption (Eq.3.45) that patterns of different landmarks are statistically independent is only acceptable if the shapes in the training images have similar orientations and scales. Indeed, if the shapes have largely differing orientations in the images, knowing the intensity pattern of one landmark contains essential information about the expected pattern to be observed at a neighboring landmark.

Several strategies for constructing the intensity descriptor (Eq. 3.44) are possible. An effective approach is presented in section 3.6.2. The corresponding intensity energy function is proposed in Section 3.6.3 and how its parameters are learnt from the training set is the topic of Section 3.6.4.

3.6.2 Intensity descriptor

The intensity descriptor (Eq. 3.44) performs two successive steps. First, a number of feature images is computed and secondly, intensity profiles are extracted from them.

Intensity profiles

A commonly used means for extracting local gray-level appearance is by sampling the image at a limited number of points around the landmark point of interest.

The standard ASM algorithm put forward by Cootes and Taylor [4] uses linear profiles centered at the landmarks and perpendicular to the object contour (or surface) in this point. In order to extract features at a point \mathbf{l}_i , the locations of the neighboring points need to be known. Consequently, such a profile is not suited for defining the descriptor we have in mind as $F(\mathbf{I}, \mathbf{l}_i)$ only depends on the image and one landmark point.

To cope with this, a spherical profile with a fixed configuration is proposed. The profile vector is extracted by sampling the intensity values in the image at $n_{\mathbf{f}}$ locations on a sphere (circle in 2D) with radius $r_{\mathbf{f}}$ centered at the landmark point. To reduce the effect of global intensity offsets between images, the profiles are normalized such that the sum of absolute values is equal to one.

In Fig. 3.9 an example configuration is shown on a slice of a computed tomography scan for modeling the image appearance of the liver. The profile consists of $n_{\mathbf{f}} = 6$ samples: two samples are put on each of the three axes.

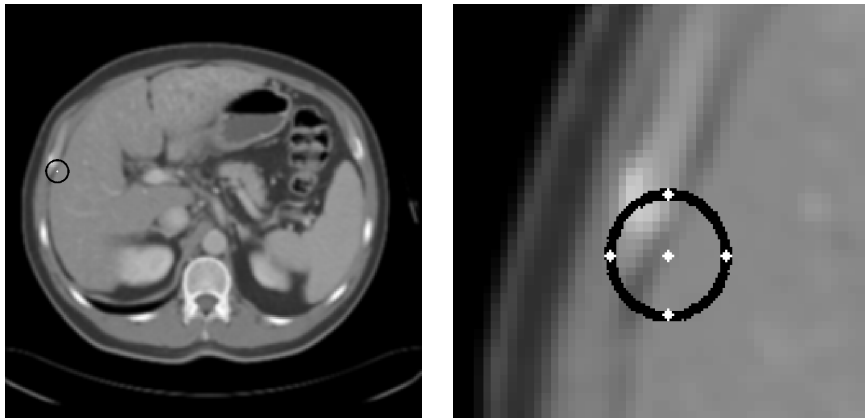


Figure 3.9: An intensity profile extracted from a computed tomography scan of the liver. The profile consists of $n_f = 6$ samples: two samples are put on each of the three axes.

Feature images

In order to construct a strong intensity model, the profile is extracted from a set of feature images $\{\mathbf{I}[1], \dots, \mathbf{I}[n_f]\}$ derived from the original image \mathbf{I} . The profile corresponding to landmark i and feature image j is denoted as $\mathbf{f}_i^{[j]}$ and the overall pattern \mathbf{f}_i is defined as the concatenation of the feature individual profiles:

$$\mathbf{f}_i = \begin{bmatrix} \mathbf{f}_i^{[1]} \\ \vdots \\ \mathbf{f}_i^{[n_f]} \end{bmatrix} \quad (3.46)$$

The approach described by van Ginneken *et al.* [8] to compute *locally orderless images* [9] is adopted here. The feature images are computed in two successive steps.

First, the original image is fed into a filter bank of multiscale Gaussian derivatives of several orders. This is motivated by the fact that the local behavior of a function $f(x)$ at a point x_0 can be described as a polynomial of some order K (Taylor expansion)

$$f(x) \approx \sum_{k=0}^K \frac{1}{k!} f^{(k)}(x_0) (x - x_0)^k \quad (3.47)$$

Hence, the coefficients $f^{(k)}(x_0)$ contain all the information about the local behavior of $f(x)$ around x_0 . For the image \mathbf{I} , this is achieved by convolving it with the derivatives of a Gaussian with scale σ (inner scale). The output of the filter bank is then a set of images resulting from multiple derivatives and scales σ .

Secondly, from each filtered image, a set of feature images is extracted by computing at each location the first few statistical moments of the local distribution of image intensities. The local histograms are computed using a Gaussian-shaped region of interest with extent α . The location dependant histograms are called *locally orderless images*. The word *orderless* refers to the fact that the spatial image structure within the field of view α is lost. The result of this procedure is a set of feature images $\mathbf{I}^{[1]}, \dots, \mathbf{I}^{[n]}$ extracted from the source image I . A detailed explanation how the LOIs are computed can be found in [86].

Fig. 3.10 shows a number of feature images computed from a chest radiograph following this approach. The filter bank consists of derivatives up to second order ($L, L_x, L_y, L_{xx}, L_{yy}, L_{xy}$) using the inner scales $\sigma = 2$ and $\sigma = 4$ pixels. The first two moments ($M = 1$ and $M = 2$) are extracted from the histograms computed with window sizes $\alpha = 2\sigma$.

A major property of this intensity descriptor $F(\mathbf{I}, \mathbf{l}_i)$ is that it is not invariant for rotations because of two reasons. First, the fact that fixed profile configurations are used and secondly, because not all the generated feature images are rotationally invariant. Two adjustments are needed for obtaining an descriptor that generates intensity patterns that do not change with a rotating image:

1. A dynamic profile configuration should be used: if the image rotates, the profile rotates with it. One way to do this is by using linear profiles sampled along the image gradient.
2. Only rotationally invariant feature images should be computed. For example the zero order derivatives of the filter bank satisfy this property.

For applications where the objects have (roughly) similar orientations and scales, this property is a clear advantage as it is more specific. So, instead of designing a descriptor that does not change with rotations, it is better to do a rough affine alignment of the images if needed. This is the approach followed in this work.

3.6.3 Intensity energy

Writing the landmark-individual intensity PDFs $p(\mathbf{f}_i)$ in the energy space:

$$p(\mathbf{f}_i) = \frac{1}{Z_{\mathbf{f}_i}} \exp\left(-\frac{1}{2}d_i(\mathbf{f}_i)\right) \quad (3.48)$$

allows to work with energies $d_i(\mathbf{f}_i)$ instead of probabilities $p(\mathbf{f}_i)$. A location \mathbf{l}_{test} in an image \mathbf{I} with a low intensity energy $d_i(F(\mathbf{I}, \mathbf{l}_{\text{test}}))$ has an intensity pattern that is similar to what is found in the training images and is likely to coincide with the true landmark location \mathbf{l}_i from an intensity point of view.

Combining Eq. 3.45 and Eq. 3.48 delivers

$$p(\mathbf{I}|\mathbf{l}) = \frac{1}{Z_{\mathbf{I}|\mathbf{l}}} \exp\left(-\frac{1}{2}\sum_{i=1}^n d_i(F(\mathbf{I}, \mathbf{l}_i))\right) \quad (3.49)$$

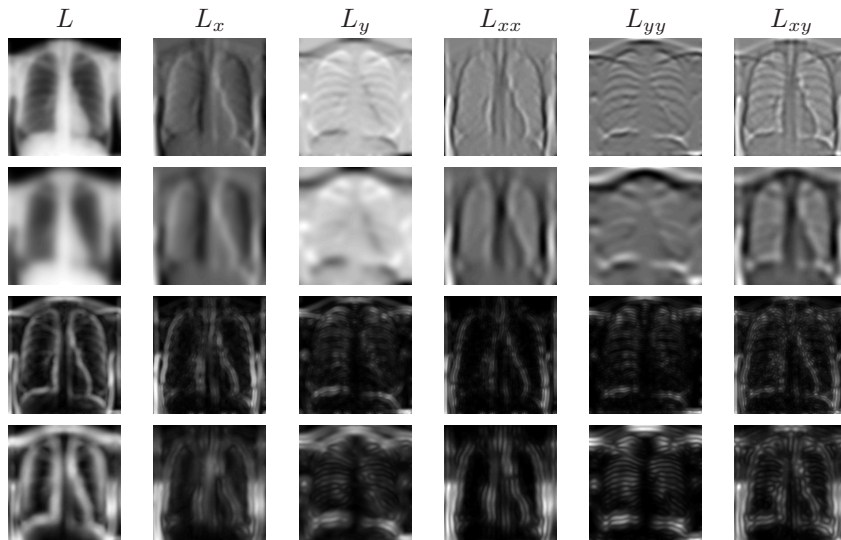


Figure 3.10: LOIs of a chest radiograph. The first ($M = 1$) and second ($M = 2$) moment images of the derivatives up to second-order ($L, L_x, L_y, L_{xx}, L_{yy}, L_{xy}$) and the scales $\sigma = 2$ and 4 pixels are shown.

with $Z_{\mathbf{I}|\mathbf{l}}$ a normalizing constant, and leads to the definition of the global intensity energy

$$E_{\mathbf{f}}(\mathbf{I}, \mathbf{l}) = \sum_{i=1}^n d_i(F(\mathbf{I}, \mathbf{l}_i)) \quad (3.50)$$

The intensity energy (Eq. 3.50) expresses how well a shape fits to the image. Minimizing this energy function with respect to \mathbf{l} , is equal to finding the most probable shape \mathbf{I} if only intensity information is considered.

3.6.4 Learning from examples

Similar as for the edge distributions, a gaussian is chosen to model the probability density of the intensity patterns \mathbf{f}_i . This results in an energy that is equal to the Mahalanobis distance between \mathbf{f}_i and its mean:

$$d_i(\mathbf{f}_i) = (\mathbf{f}_i - \boldsymbol{\mu}_{\mathbf{f}_i})^T \boldsymbol{\Sigma}_{\mathbf{f}_i}^{-1} (\mathbf{f}_i - \boldsymbol{\mu}_{\mathbf{f}_i}) \quad (3.51)$$

In general an unacceptable large training set is needed to estimate the distribution parameters $\boldsymbol{\mu}_{\mathbf{f}_i} \in \mathbb{R}^{n_{\mathbf{f}_i}}$ and $\boldsymbol{\Sigma}_{\mathbf{f}_i} \in \mathbb{R}^{n_{\mathbf{f}_i} \times n_{\mathbf{f}_i}}$. To cope with this, the

$n_{\mathbf{I}}$ sub-profiles $\mathbf{f}_i^{[j]}$ are modeled as if they were independent:

$$\Sigma_{\mathbf{f}_i} = \begin{bmatrix} \Sigma_{\mathbf{f}_i^{[1]}} & 0 & \dots & 0 \\ 0 & \Sigma_{\mathbf{f}_i^{[2]}} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \Sigma_{\mathbf{f}_i^{[n_{\mathbf{I}]}}}] \end{bmatrix} \quad (3.52)$$

and $\boldsymbol{\mu}_{\mathbf{f}_i}^{\mathbf{T}} = [\boldsymbol{\mu}_{\mathbf{f}_i^{[1]}}^{\mathbf{T}} \dots \boldsymbol{\mu}_{\mathbf{f}_i^{[n_{\mathbf{I}]}}}]^{\mathbf{T}}$. The intensity energy of Eq. 3.51 becomes

$$d_i(\mathbf{f}_i) = \sum_{j=1}^{n_{\mathbf{I}}} (\mathbf{f}_i^{[j]} - \boldsymbol{\mu}_{\mathbf{f}_i^{[j]}})^{\mathbf{T}} \Sigma_{\mathbf{f}_i^{[j]}}^{-1} (\mathbf{f}_i^{[j]} - \boldsymbol{\mu}_{\mathbf{f}_i^{[j]}}) \quad (3.53)$$

with the distribution parameters $\boldsymbol{\mu}_{\mathbf{f}_i^{[j]}}$ and $\Sigma_{\mathbf{f}_i^{[j]}}$ estimated from the training samples for each vertex i and for each LOI j . The training samples are obtained by extracting the intensity patterns (Eq. 3.44) from the training images at the landmark positions \mathbf{l}_i of the corresponding shapes.

The energy in Eq. 3.53 is only exact if the independency assumption (Eq. 3.52) is true. In practice, this is not the case. It can be proven that the energy function 3.53 is always an overestimation of the true value if the independency assumption is violated. One way to obtain a more realistic probability, is to rescale the energy:

$$d_i(\mathbf{f}_i) = \frac{1}{\rho} \sum_{j=1}^{n_{\mathbf{I}}} (\mathbf{f}_i^{[j]} - \boldsymbol{\mu}_{\mathbf{f}_i^{[j]}})^{\mathbf{T}} \Sigma_{\mathbf{f}_i^{[j]}}^{-1} (\mathbf{f}_i^{[j]} - \boldsymbol{\mu}_{\mathbf{f}_i^{[j]}}) \quad (3.54)$$

with $\rho \geq 1$ a parameter to be estimated. The amount of overestimation is related to how strong the $n_{\mathbf{I}}$ different sub-profiles are correlated. The more the non-diagonal elements of Eq 3.52 differ from zero the higher ρ must be chosen. If the feature images would be uncorrelated, then Eq. 3.52 is exact and the intensity energy computed as 3.53 is valid or $\rho = 1$. In the opposite case (if all profiles are 100% correlated), the energy overestimates the real energy exactly $n_{\mathbf{I}}$ times: $\rho = n_{\mathbf{I}}$.

This is illustrated in Fig. 3.11. A radiograph of the hand is scanned with the intensity energy (Eq. 3.53) of the landmarks shown in the templates (a). Fig. 3.11(b) shows the corresponding probability densities (Eq. 3.48). The points in the image with the highest values (whitest regions) indicate where the landmark is expected to lie. As only one white spot is visible, the intensity energy as computed in Eq. 3.53 puts a too high restriction on other possible locations. If the same distributions are computed with a corrected energy (Eq. 3.54), a more truthful result is obtained as shown in Fig. 3.11(c). A procedure to obtain an appropriate ρ -value is given in Chapter 4.

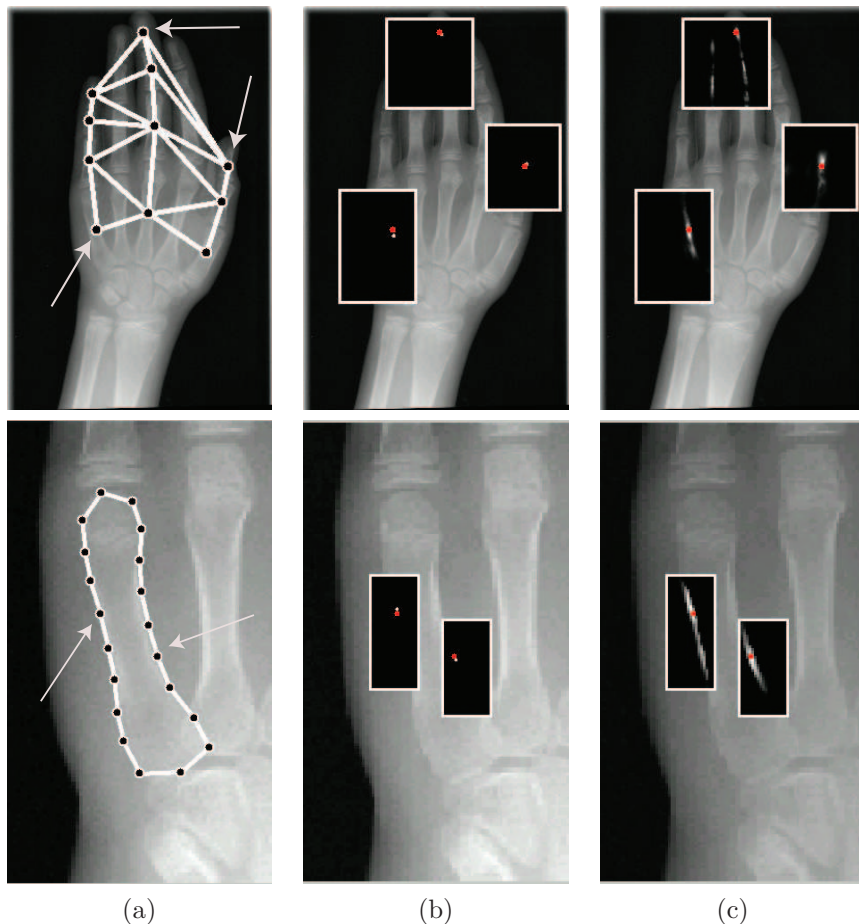


Figure 3.11: A radiograph of the hand is scanned with the intensity energy (Eq. 3.53) of the landmarks shown in the templates (a). The corresponding probability densities (Eq. 3.48) are shown in (b). It seems that the probability values are peaked as only one white spot is visible. A less peaked Gibbs distribution leads to a more realistic distribution as shown in (c).

3.7 Combined model

When an unseen image has to be segmented, a shape will be searched that properly fits the image data according to the appearance model. The shape model at the other hand is needed to guarantee that only plausible shapes are generated. The statistical approach of the segmentation task combines the shape and appearance information in a natural way:

$$p(\mathbf{l}|\mathbf{I}) \propto p(\mathbf{l})p(\mathbf{I}|\mathbf{l}) \quad (3.55)$$

with $p(\mathbf{l})$ and $p(\mathbf{I}|\mathbf{l})$ respectively the shape and intensity PDFs:

$$p(\mathbf{l}) = \frac{1}{Z_1} \exp\left(-\frac{1}{2}E_1(\mathbf{l})\right) \quad (3.56)$$

$$p(\mathbf{I}|\mathbf{l}) = \frac{1}{Z_{\mathbf{I}|\mathbf{l}}} \exp\left(-\frac{1}{2}E_f(\mathbf{I}, \mathbf{l})\right) \quad (3.57)$$

and with $E_1(\mathbf{l})$ and $E_f(\mathbf{I}, \mathbf{l})$ respectively the shape energy (Eq. 3.38) and intensity energy (Eq. 3.50).

Object energy

Eq. 3.55 becomes (using Eq. 3.56 and Eq. 3.57)

$$p(\mathbf{l}|\mathbf{I}) = \frac{1}{Z} \exp\left(-\frac{1}{2}E(\mathbf{l}, \mathbf{I})\right) \quad (3.58)$$

with Z a constant with respect to \mathbf{l} and with $E(\mathbf{l}, \mathbf{I})$ called the *object energy*:

$$E(\mathbf{l}, \mathbf{I}) = \sum_{i \in \mathcal{V}} d_i(F(\mathbf{I}, \mathbf{l}_i)) + \frac{n-1}{2t} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \quad (3.59)$$

or

$$E(\mathbf{l}, \mathbf{I}) = \sum_{i \in \mathcal{V}} \left(d_i(F(\mathbf{I}, \mathbf{l}_i)) + \frac{n-1}{2t} \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \right). \quad (3.60)$$

A low energy suggests that the shape fits both the image data and the shape prior. The intensity term in Eq. 3.59 drives the shape to the correct image features whereas the second term controls the shape.

Markov object energy

From Eq. 3.58 and Eq. 3.59 it follows that the shape \mathbf{l} for a given image \mathbf{I} behaves as an MRF on \mathcal{V} with respect to \mathcal{N} . The Gibbs energy consists of first order intensity potentials and second order shape potentials.

The probability of a landmark location \mathbf{l}_i given the image \mathbf{I} and also given the locations of the neighbors $\mathbf{l}_{\mathcal{N}_i}$ is computed using Eq. 3.59 and Eq. 3.12

$$p(\mathbf{l}_i|\mathbf{I}, \mathbf{l}_{\mathcal{N}_i}) = K(\mathbf{I}, \mathbf{l}_{\mathcal{N}_i}) \exp\left(-\frac{1}{2} \left(d_i(F(\mathbf{I}, \mathbf{l}_i)) + \frac{n-1}{t} \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \right)\right) \quad (3.61)$$

with K a constant with respect to \mathbf{l}_i .

Hence, the conditional energy is a sum of the intensity energy of \mathbf{l}_i and the conditional shape energy of \mathbf{l}_i given its neighbors $\mathbf{l}_{\mathcal{N}_i}$ (Eq. 3.39). This is illustrated by Fig. 3.12 for the probability distribution of specific landmarks of the hand shape (a). The intensity PDF $p(\mathbf{f}_i)$ (b) with $\mathbf{f}_i = F(\mathbf{I}, \mathbf{l}_i)$ expresses where \mathbf{l}_i is expected to lie as only intensity information is used, (c) shows where \mathbf{l}_i is expected if the location of the neighbors $\mathbf{l}_{\mathcal{N}_i}$ is known and (d) combines the knowledge of \mathbf{I} and $\mathbf{l}_{\mathcal{N}_i}$ to estimate \mathbf{l}_i .

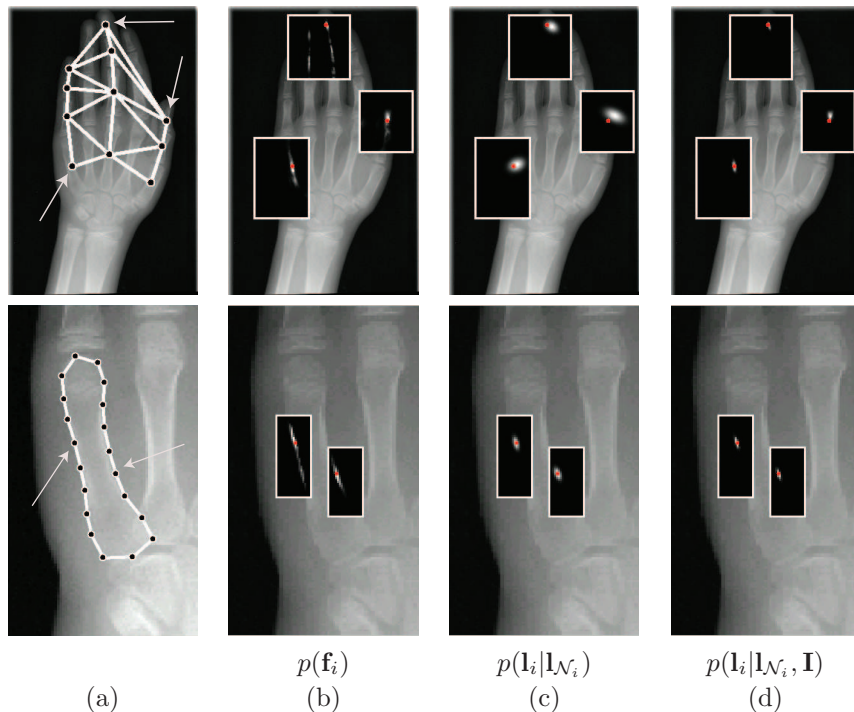


Figure 3.12: The probability of a landmark location \mathbf{l}_i given the locations of the neighbors $\mathbf{l}_{\mathcal{N}_i}$ and the image \mathbf{I} for the hand shapes (a). The intensity PDF $p(\mathbf{f}_i)$ (b) with $\mathbf{f}_i = F(\mathbf{I}, \mathbf{l}_i)$ expresses where \mathbf{l}_i is expected to lie as only intensity information is used, (c) shows where \mathbf{l}_i is expected if the location of the neighbors $\mathbf{l}_{\mathcal{N}_i}$ is known and (d) combines the knowledge of \mathbf{I} and $\mathbf{l}_{\mathcal{N}_i}$ to estimate \mathbf{l}_i .

3.8 Conclusion

This chapter describes how prior information about local properties of an object to be segmented is captured. Towards finding the most likely segmented object in an unseen image, prior knowledge about the object geometry (shape model) and about how the object appears in the image (intensity model) is needed. Both models are learnt from examples.

The object to be segmented is represented as a graph that consists of a number of landmark points defined on the shape and a number of connections between neighboring landmark points that express the shape correlations. Such an approach allows to extract local statistical properties, both shape and appearance related.

As the proposed shape model assumes that a landmark point only interacts with its neighbors, the shape behaves as a Markov random field. Writing the shape probability density as a Gibbs function allows to work with shape ener-

gies rather than probabilities. The shape energy is expressed as a sum of local energies that only depend on couples of neighboring landmark points.

The appearance model assumes that the object controls the local image structure around each landmark individually. The local intensity patterns are constructed by extracting profiles from a number of feature images computed from the original image. By estimating the statistical distributions of the gray-level patterns and writing them as Gibbs functions, a global intensity energy can be calculated. This global energy consists of landmark-individual appearance energies and expresses the likelihood of the observed image patterns in the neighborhood of the shape.

Both models are merged into one object model. A corresponding object energy is defined that consists of first order intensity potentials and second order shape potentials. The intensity potentials drive the shape towards the correct image features whereas the shape potentials control the shape geometry.

Chapter 4

Model fitting

This chapter explains how the object model that was introduced in Chapter 3, is used to find the location of that object in an unseen image. Thanks to the statistical nature of the model, the *model fitting* could be formulated from a probability point of view: given an image, and also given a model that describes how the object looks like, find its most likely location. Prior to optimizing the energy function, the segmentation task is first converted into a discrete optimization problem (Section 4.1). A number of optimization algorithm are given in Section 4.2 and some important implementation issues are reported in Section 4.3.

4.1 Objective function

4.1.1 Continuous optimization problem

The segmentation task is considered as a maximum a posteriori problem (MAP):

$$\mathbf{I}^* = \arg \max_{\mathbf{I}} p(\mathbf{I}|\mathbf{I}) \quad (4.1)$$

The optimization problem 4.1 is equivalent to minimizing the object energy (Eq. 3.59):

$$\mathbf{I}^* = \arg \min_{\mathbf{I}} E(\mathbf{I}, \mathbf{I}) \quad (4.2)$$

with

$$E(\mathbf{I}, \mathbf{I}) = \sum_{i \in \mathcal{V}} d_i(F(\mathbf{I}, \mathbf{l}_i)) + \frac{n-1}{2t} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \quad (4.3)$$

The difference in problem formulation between this approach and the ASM methods (Section 2.4.1) is that the energy to be optimized incorporates both shape and image information. The ASM search method implicitly optimizes an energy that only contains intensity information. Shape knowledge is used to constrain the solution \mathbf{I} to lie within a set of plausible shapes:

$$\mathbf{I}^* = \arg \min_{\mathbf{I}=\mu_1+\Phi\mathbf{b}} E_f(\mathbf{I}, \mathbf{I}) \quad (4.4)$$

Table 4.1: Methodological properties of ASM and MISCG.

Property	ASM	MISCG
Cost function	intensity	intensity and shape
Optimization	constrained by shape	unconstrained
Shape model	global	local

Another major difference is the nature of the shape model. The proposed method uses multiple localized shape models whereas ASM applies one global shape model. A set of localized shape models is weaker than one global model in the sense that it enables unrealistic shapes, but it is more flexible and thus enables an accurate fit to the image. These properties are summarized in Table 4.1. The proposed method will in this text be referred to as the MISCG method. The acronym stands for *Minimal Intensity and Shape Cost Graph*.

4.1.2 Discrete optimization problem

As the objective function 4.3 is non-convex and has a large number of independent variables, conventional optimization techniques will steer the shape into meaningless local optima. To cope with this, prior to optimization, the problem is converted into a discrete optimization problem. The search space is discretized by searching a number of appropriate candidate locations for each landmark \mathbf{l}_i . To accomplish this, the image (or a relevant part of it) is scanned with the intensity energy function $d_i(F(\mathbf{I}, \mathbf{l}_i))$ (Eq. 3.54) as illustrated in Fig 4.1. For each landmark, the m lowest-energy-locations are selected resulting in the *candidate set*

$$\mathcal{C} = \{\{\mathbf{l}_{ik}, \mathbf{f}_{ik}\}_{k=1}^m\}_{i=1}^n \quad (4.5)$$

At this point, the segmentation task is converted into a labeling problem \mathbf{x} that selects one candidate for each landmark:

$$\mathbf{x} : \mathcal{V} = \{1, \dots, n\} \rightarrow \mathbb{X} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} \quad (4.6)$$

A solution $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$ with $\mathbf{x}_i \in \mathbb{X}$ is called a *labeling* and the solution space is noted as \mathbb{X}^n . According to the labeling, the landmark locations and intensity features can be written as

$$\mathbf{l}_i = \sum_{a=1}^m x_{ia} \mathbf{l}_{ia} \quad (4.7)$$

$$\mathbf{f}_i = \sum_{a=1}^m x_{ia} \mathbf{f}_{ia} \quad (4.8)$$

	continuous	discrete
random field	\mathbf{l}	\mathbf{x}
probability	$p(\mathbf{l} \mathbf{I})$	$P(\mathbf{x} \mathcal{C})$
energy	$E(\mathbf{l}, \mathbf{I})$	$E(\mathbf{x}, \mathcal{C})$
optimization	$\mathbf{l}^* = \arg \min_{\mathbf{l}} E(\mathbf{l}, \mathbf{I})$	$\mathbf{x}^* = \arg \min_{\mathbf{x}} E(\mathbf{x}, \mathcal{C})$

Table 4.2: By discretizing the search space of the shape \mathbf{l} , the segmentation problem becomes a labeling problem \mathbf{x} .

with x_{ij} the j -th element of \mathbf{x}_i . By the defined discretization (Table 4.2), the segmentation task is converted into the discrete optimization problem

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} E(\mathbf{x}, \mathcal{C}) \quad (4.9)$$

with

$$E(\mathbf{x}, \mathcal{C}) = \sum_{i \in \mathcal{V}} \sum_{a=1}^m x_{ia} \left(d_i(\mathbf{f}_{ia}) + \frac{n-1}{2t} \sum_{j \in \mathcal{N}_i} \sum_{b=1}^m x_{jb} d_{ij}(\mathbf{l}_{ia}, \mathbf{l}_{jb}) \right) \quad (4.10)$$

As the landmark locations \mathbf{l}_i and features \mathbf{f}_i can be expressed in terms of the candidates \mathcal{C} and the labels \mathbf{x} (Eq. 4.7-4.8), the energies $d_i(\mathbf{f}_i)$ and $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$ can also be noted as functions of \mathbf{x} and \mathcal{C} . With this notation, the energy function becomes

$$E(\mathbf{x}, \mathcal{C}) = \sum_{i \in \mathcal{V}} d_i(\mathbf{x}_i, \mathcal{C}_i) + \frac{n-1}{t} \sum_{\{i,j\} \in \mathcal{E}} d_{ij}(\mathbf{x}_i, \mathbf{x}_j, \mathcal{C}_i, \mathcal{C}_j) \quad (4.11)$$

The total energy consists of first order (intensity) energies $d_i(\mathbf{x}_i)$ assigned to the nodes in the graph, and second order (shape) energies $d_{ij}(\mathbf{x}_i, \mathbf{x}_j)$ assigned to the edges. Both expressions (Eq. 4.10 and Eq. 4.11) will be used in the following section.

4.2 Optimization techniques

Three discrete optimization algorithms are proposed to find a candidate for each landmark that properly fits to the image data and leads to a plausible shape configuration (Eq. 4.10 and Eq. 4.11).

4.2.1 Dynamic programming

The term *dynamic programming* (DP) was first used by Richard Bellman [10] to describe the process of solving problems where one needs to find the best decisions one after the other. DP algorithms are only applicable for problems that have the *optimal substructure* property. This means that optimal

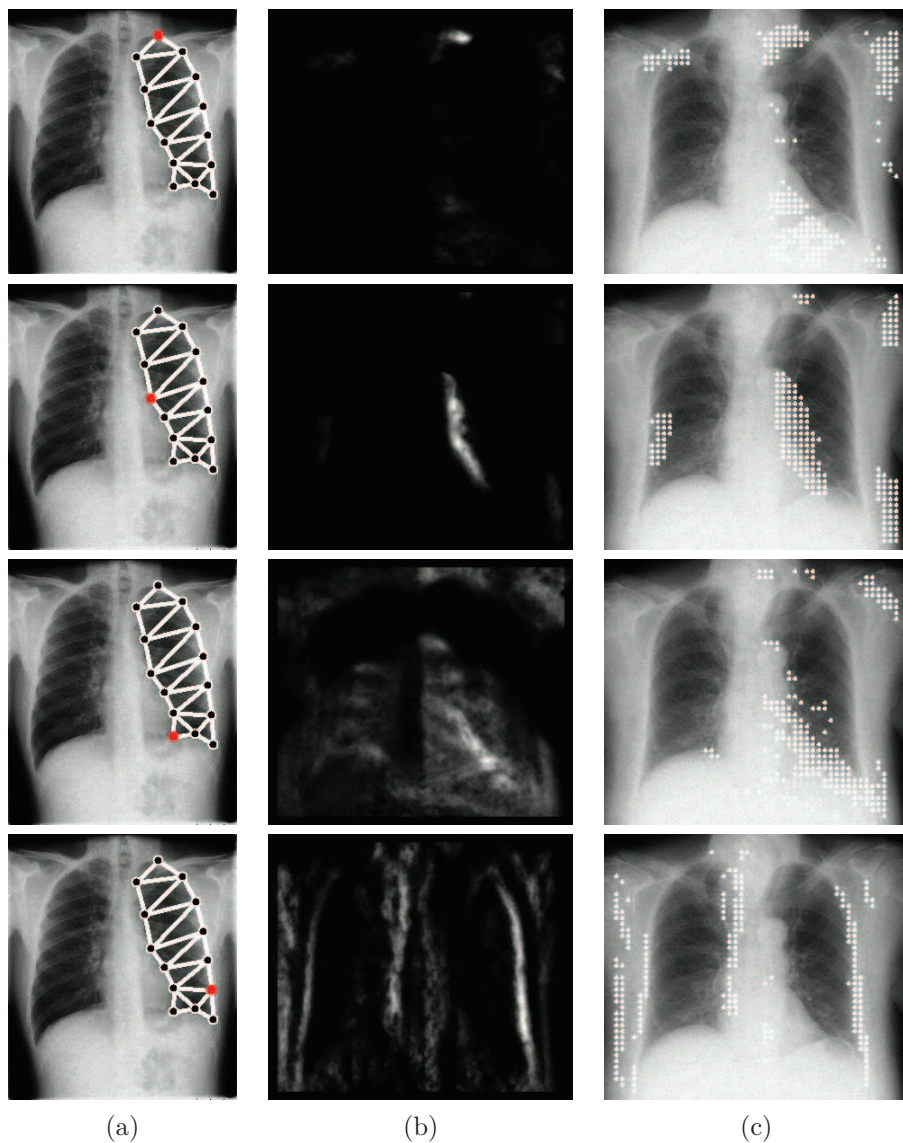


Figure 4.1: The search problem is converted into a discrete optimization problem by discretizing the search space. The intensity energy function $d_i(F(\mathbf{I}, \mathbf{l}_i))$ is used to detect a number of candidate locations for each landmark individually. The landmarks of interest are indicated in (a) and the corresponding probabilities are shown in (b). The best 200 candidates of a search grid that covers the complete image area are selected in (c).

solutions of subproblems can be used to find the optimal solution of the overall problem. DP has been applied for many image processing tasks (e.g. [20, 24, 25, 29, 73, 74]). Here a DP algorithm is described that finds the minimal cost E^* and corresponding labels \mathbf{x}^* of the energy function 4.11. First, the DP optimization scheme is given in its most general form, such that it can be applied to any graph configuration. Afterwards implementations for a number of specific graph configurations will be given.

General formulation

To benefit from the optimal substructure property, the graph is reduced landmark by landmark. The original graph consisting of n landmarks, is reduced to one with $n - 1$ landmarks by eliminating one vertex and all the corresponding edges. If we can find the solution for the reduced graph, the optimal labeling of the complete graph can be computed easily. Without loss of generality, it is assumed that the graph is eliminated using the order 1, 2, ..., n .

This is mathematically formulated as follows. First, the minimal energy E^* is expressed as a function of the (unknown) optimal labeling \mathbf{x}^* :

$$E^* = E(\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_n^*) \quad (4.12)$$

To eliminate the first vertex, the dependency on \mathbf{x}_1^* in Eq. 4.12 needs to be removed. Therefore, Eq. 4.12 is reformulated as

$$E^* = \min_{\mathbf{x}_1} E(\mathbf{x}_1, \mathbf{x}_2^*, \dots, \mathbf{x}_n^*) \quad (4.13)$$

$$= \min_{\mathbf{x}_1} \left[d_1(\mathbf{x}_1) + \frac{n-1}{t} \sum_{j \in \mathcal{N}_1} d_{1j}(\mathbf{x}_1, \mathbf{x}_j^*) + E_1(\mathbf{x}_2^*, \dots, \mathbf{x}_n^*) \right] \quad (4.14)$$

$$= \min_{\mathbf{x}_1} \left[d_1(\mathbf{x}_1) + \frac{n-1}{t} \sum_{j \in \mathcal{N}_1} d_{1j}(\mathbf{x}_1, \mathbf{x}_j^*) \right] + E_1(\mathbf{x}_2^*, \dots, \mathbf{x}_n^*) \quad (4.15)$$

$$= Q_1(\mathbf{x}_2^*, \dots, \mathbf{x}_n^*) + E_1(\mathbf{x}_2^*, \dots, \mathbf{x}_n^*) \quad (4.16)$$

Q_1 is referred to as the elimination energy, as it is the sum of the energies of the eliminated parts and E_1 the energy of the remaining vertices and edges. In the first elimination step, the elimination energy is only function of $\mathbf{x}_{\mathcal{N}_1}^*$:

$$Q_1(\mathbf{x}_2^*, \dots, \mathbf{x}_n^*) = q_{\mathcal{N}_1}(\mathbf{x}_{\mathcal{N}_1}^*) \quad (4.17)$$

To compute $q_{\mathcal{N}_1}$ for a specific $\mathbf{x}_{\mathcal{N}_1}^*$, the m possibilities for \mathbf{x}_1 need to be compared. Once $q_{\mathcal{N}_1}$ is calculated for every possible $\mathbf{x}_{\mathcal{N}_1}^*$ and stored in a multi-dimensional array, the dependency of E^* on \mathbf{x}_1^* is removed (Eq. 4.16). At this point, the energy E^* equals the sum of the energies d_i of the remaining vertices and the energies d_{ij} of the remaining edges plus a known energy term $q_{\mathcal{N}_1}$ that depends on the vertices in \mathcal{N}_1 . During optimization, one also keeps

track of the optimal label of the eliminated vertex:

$$\mathbf{x}_1^* = R_1(\mathbf{x}_{\mathcal{N}_1}^*) = \arg \min_{\mathbf{x}_1} \left[d_1(\mathbf{x}_1) + \frac{n-1}{t} \sum_{j \in \mathcal{N}_1} d_{1j}(\mathbf{x}_1, \mathbf{x}_j^*) \right] \quad (4.18)$$

The procedure continues with the elimination of vertex 2. Similar as in Eq. 4.13-4.16, one obtains:

$$E^* = Q_2(\mathbf{x}_3^*, \dots, \mathbf{x}_n^*) + E_2(\mathbf{x}_3^*, \dots, \mathbf{x}_n^*) \quad (4.19)$$

with Q_2 the minimal energy to eliminate vertices 1 and 2:

$$Q_2(\mathbf{x}_3^*, \dots, \mathbf{x}_n^*) = \min_{\mathbf{x}_2} \left[d_2(\mathbf{x}_2) + \frac{n-1}{t} \sum_{j \in \mathcal{N}'_2} d_{2j}(\mathbf{x}_2, \mathbf{x}_j^*) + Q_1(\mathbf{x}_2, \mathbf{x}_3^*, \dots, \mathbf{x}_n^*) \right] \quad (4.20)$$

where \mathcal{N}'_2 contains the neighbors of vertex 2 except vertex 1. Obviously, two different situations exist:

1. If Q_1 does not depend on vertex 2 (if vertex 2 is not a neighbor of vertex 1), then Q_2 splits up in a sum of two functions: $Q_2 = q_{\mathcal{N}_1}(\mathbf{x}_{\mathcal{N}_1}^*) + q_{\mathcal{N}_2}(\mathbf{x}_{\mathcal{N}_2}^*)$. The first one corresponds to the elimination of vertex 1 as computed in the previous step. The second one corresponds to the elimination of vertex 2 and needs to be computed for every possible $\mathbf{x}_{\mathcal{N}_2}^*$.
2. When at the other hand, vertex 2 is a neighbor, then $Q_2 = q_{\mathcal{A}}(\mathbf{x}_{\mathcal{A}}^*)$ consists of only one subfunction that has a value for every possible $\mathbf{x}_{\mathcal{A}}^*$ with $\mathcal{A} = (\mathcal{N}_1 \cup \mathcal{N}_2) / \{1, 2\}$.

In general, after eliminating landmark i one has

$$E^* = Q_i(\mathbf{x}_{i+1}^*, \dots, \mathbf{x}_n^*) + E_i(\mathbf{x}_{i+1}^*, \dots, \mathbf{x}_n^*) \quad (4.21)$$

and the elimination energy Q_i until and including vertex i follows from the update equation

$$Q_i(\mathbf{x}_{i+1}^*, \dots, \mathbf{x}_n^*) = \min_{\mathbf{x}_i} \left[d_i(\mathbf{x}_i) + \frac{n-1}{t} \sum_{j \in \mathcal{N}'_i} d_{ij}(\mathbf{x}_i, \mathbf{x}_j^*) + Q_{i-1}(\mathbf{x}_i, \mathbf{x}_{i+1}^*, \dots, \mathbf{x}_n^*) \right] \quad (4.22)$$

with $\mathcal{N}'_i = \{j \in \mathcal{N}_i | j > i\}$. If Q_{i-1} can be written as a sum of subfunctions that depend on limited sets of vertices, this also counts for Q_i :

$$Q_i(\mathbf{x}_{i+1}^*, \dots, \mathbf{x}_n^*) = \sum_k q_{\mathcal{A}_{i,k}}(\mathbf{x}_{\mathcal{A}_{i,k}}^*) \quad (4.23)$$

During optimization, one has to keep track of the subfunctions $q_{\mathcal{A}_{i,k}}$ for all possible values of $\mathbf{x}_{\mathcal{A}_{i,k}}^*$. Also the corresponding optimal labels $\mathbf{x}_{i,k}^*$ are stored

in multi-dimensional arrays R_i . The elimination process continues till only one landmark is left:

$$E^* = Q_{n-1}(\mathbf{x}_n^*) + E_{n-1}(\mathbf{x}_n^*) \quad (4.24)$$

The optimal label \mathbf{x}_n^* and minimal energy E^* are obtained from Eq. 4.20 by comparing the m possibilities of \mathbf{x}_n^* . Finally, a tracing back procedure using the stored arrays R_i delivers the desired labels $\mathbf{x}_{n-1}^*, \mathbf{x}_{n-2}^*, \dots, \mathbf{x}_1^*$.

The DP procedure can be applied to any graph, but the computational complexity highly depends on the configuration of the graph and the landmark ordering. During optimization, in each elimination step, a cost $q_{\mathcal{A}}(\mathbf{x}_{\mathcal{A}}^*)$ is computed for every possible $\mathbf{x}_{\mathcal{A}}^*$. As the cost for a specific $\mathbf{x}_{\mathcal{A}}^*$ is found as the minimum of m possible values, the computational complexity of the elimination step is $O(m^{|\mathcal{A}|+1})$ with $|\mathcal{A}|$ the number of vertices in \mathcal{A} . For the graph in Fig. 4.2(a), the elimination costs only depend on one vertex and hence, optimization runs in $O(nm^2)$. The examples shown in (b) and (c) have elimination steps that run in respectively $O(m^3)$ and $O(m^4)$.

From a practical point of view, DP is only applicable if $|\mathcal{A}|$ is small enough. In Sections 4.2.2 and 4.2.3 optimization methods will be proposed that optimize in $O(m^2)$ irrespective of the complexity of the graph.

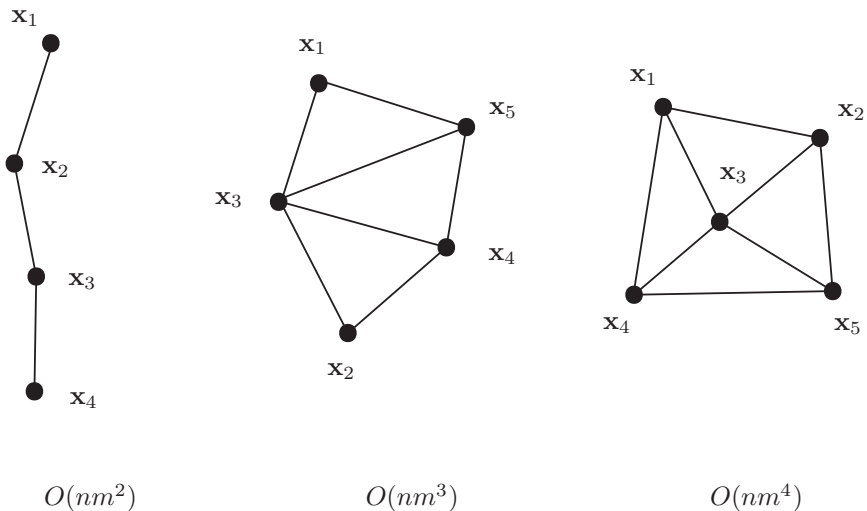


Figure 4.2: Dynamic programming can be used to find the optimal position for each landmark. The computational complexity depends on the configuration of the graph.

Example

Consider the example graph shown in Fig. 4.2(b). Applying Eq. 4.22 delivers the following elimination scheme:

$$\begin{aligned}
 E^* &= E(\mathbf{x}_1^*, \mathbf{x}_2^*, \mathbf{x}_3^*, \mathbf{x}_4^*, \mathbf{x}_5^*) \\
 &= E_1(\mathbf{x}_2^*, \mathbf{x}_3^*, \mathbf{x}_4^*, \mathbf{x}_5^*) + q_{3,5}(\mathbf{x}_3^*, \mathbf{x}_5^*) \\
 &= E_2(\mathbf{x}_3^*, \mathbf{x}_4^*, \mathbf{x}_5^*) + q_{3,5}(\mathbf{x}_3^*, \mathbf{x}_5^*) + q_{3,4}(\mathbf{x}_3^*, \mathbf{x}_4^*) \\
 &= E_3(\mathbf{x}_4^*, \mathbf{x}_5^*) + q_{4,5}(\mathbf{x}_4^*, \mathbf{x}_5^*) \\
 &= E_4(\mathbf{x}_5^*) + q_5(\mathbf{x}_5^*)
 \end{aligned}$$

with the subfunctions $q_{A_i,k}$ computed as:

$$\begin{aligned}
 q_{3,5}(\mathbf{x}_3^*, \mathbf{x}_5^*) &= \min_{\mathbf{x}_1} \left[d_1(\mathbf{x}_1) + \frac{4}{7}d_{1,3}(\mathbf{x}_1, \mathbf{x}_3^*) + \frac{4}{7}d_{1,5}(\mathbf{x}_1, \mathbf{x}_5^*) \right], \\
 q_{3,4}(\mathbf{x}_3^*, \mathbf{x}_4^*) &= \min_{\mathbf{x}_2} \left[d_2(\mathbf{x}_2) + \frac{4}{7}d_{2,3}(\mathbf{x}_2, \mathbf{x}_3^*) + \frac{4}{7}d_{2,4}(\mathbf{x}_2, \mathbf{x}_4^*) \right], \\
 q_{4,5}(\mathbf{x}_4^*, \mathbf{x}_5^*) &= \min_{\mathbf{x}_3} \left[d_3(\mathbf{x}_3) + \frac{4}{7}d_{3,4}(\mathbf{x}_3, \mathbf{x}_4^*) + \frac{4}{7}d_{3,5}(\mathbf{x}_3, \mathbf{x}_5^*) \right. \\
 &\quad \left. + q_{3,4}(\mathbf{x}_3, \mathbf{x}_4^*) + q_{3,5}(\mathbf{x}_3, \mathbf{x}_5^*) \right], \\
 q_5(\mathbf{x}_5^*) &= \min_{\mathbf{x}_4} \left[d_4(\mathbf{x}_4) + \frac{4}{7}d_{4,5}(\mathbf{x}_4, \mathbf{x}_5^*) + q_{4,5}(\mathbf{x}_4, \mathbf{x}_5^*) \right].
 \end{aligned}$$

Open contour

An open contour can be described as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with

$$\mathcal{V} = \{1, \dots, n\} \quad (4.25)$$

$$\mathcal{E} = \{\{i, i+1\}\}_{i=1}^{n-1} \quad (4.26)$$

The landmarks are labeled 1, 2, ..., n when running along the contour from one end to the other end. The update equation (Eq. 4.22) becomes:

$$Q_i(\mathbf{x}_{i+1}^*) = \min_{\mathbf{x}_i} [d_i(\mathbf{x}_i) + d_{i,i+1}(\mathbf{x}_i, \mathbf{x}_{i+1}^*) + Q_{i-1}(\mathbf{x}_i)] \quad (4.27)$$

As the elimination cost only depends on one remaining landmark, the DP algorithm (listed in Algorithm 4.4.1) has computational complexity $O(nm^2)$.

Closed contour

A closed contour is represented as the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with

$$\mathcal{V} = \{1, \dots, n\} \quad (4.28)$$

$$\mathcal{E} = \{\{i, i+1\}\}_{i=1}^{n-1} \cup \{1, n\} \quad (4.29)$$

The landmarks are ordered from 1 to n when running along the contour and the connection $\{1, n\}$ closes the path. As two edges need to be removed in the first elimination step, the elimination cost will always depend on two remaining landmarks and consequently, the optimization (listed in Algorithm 4.4.2) runs in $O(nm^3)$.

An alternative DP approach finds a solution for a closed contour in $O(nm^2)$ but without the certainty that the global optimum is found. First, the cheapest paths are searched from landmark 1 to all the possible choices for landmark n as if it was an open contour by ignoring the connection $\{1, n\}$. This is obtained by running algorithm 4.4.1 and tracing back the path for each possible \mathbf{x}_n . This way, the cheapest paths for all possible choices for the last landmark are found. Subsequently, each of these n paths are closed by adding the cost for the remaining connection $\{1, n\}$. The path with the cheapest cost is then selected. In this text, the algorithm will be referred to as dynamic programming closed contour (DPCC) (dynamic programming closed contour).

Exact energy

The DP scheme was applied to the energy function 4.11 where third (and higher) order potentials are approximated using only second order potential functions. However, without going into details, the elimination principle is also applicable to the exact energy function as well.

4.2.2 Mean field annealing

Mean field annealing (MFA) [7] is a well known optimization technique in the world of Markov random fields. It searches for a minimal Gibbs energy, not by minimizing the energy function directly, but by estimating the mean field for a decreasing Gibbs temperature.

Labeling as an MRF

In Chapter 3 it was stated that the localization of an object \mathbf{l} in a given image \mathbf{I} behaves as an MRF on $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Analogous, the labeling \mathbf{x} for a given set of candidates \mathcal{C} acts as an MRF with distribution

$$P(\mathbf{x}) = \frac{1}{Z_{\mathbf{x}}} e^{-\frac{1}{T} E(\mathbf{x}, \mathcal{C})} \quad (4.30)$$

with $E(\mathbf{x}, \mathcal{C})$ computed as in Eq. 4.10. To keep the notations simple, we write $P(\mathbf{x})$ instead of $P(\mathbf{x}|\mathcal{C})$. In the following formulas, the candidates \mathcal{C} are assumed to be given implicitly.

As the energy is a sum of local (intensity) energies that depend on one vertex and local (shape) energies that are function of two neighboring vertices, the random field \mathbf{x} given \mathcal{C} behaves as an MRF on $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Markov distribution

The conditional MRF distributions can be computed by applying Eq. 3.12 on the energy 4.10:

$$P(\mathbf{x}_i | \mathbf{x}_{\mathcal{N}_i}) = \frac{\exp\left(-\frac{1}{T}V_i(\mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i})\right)}{\sum_{\mathbf{x}'_i \in \mathbb{X}} \exp\left(-\frac{1}{T}V_i(\mathbf{x}'_i, \mathbf{x}_{\mathcal{N}_i})\right)} \quad (4.31)$$

with V_i the sum of the local energies that depend on \mathbf{x}_i :

$$V_i(\mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i}) = \sum_{a=1}^m x_{ia} d_i(\mathbf{f}_{ia}) + \frac{n-1}{t} \sum_{a=1}^m x_{ia} \sum_{j \in \mathcal{N}_i} \sum_{b=1}^m x_{jb} d_{ij}(\mathbf{l}_{ia}, \mathbf{l}_{jb}) \quad (4.32)$$

Mean field

The mean of the random field \mathbf{x} is

$$\langle \mathbf{x} \rangle = \sum_{\mathbf{x} \in \mathbb{X}^n} \mathbf{x} P(\mathbf{x}) \quad (4.33)$$

The field values x_{ij} can only take the binary values 0 and 1, whereas the mean field values $\langle x_{ij} \rangle$ are real numbers in $[0, 1]$. The mean changes with T as follows:

$$\lim_{T \rightarrow 0^+} \langle \mathbf{x} \rangle_T = \mathbf{x}^* \quad (4.34)$$

Eq. 4.34 implies that instead of minimizing the energy $E(\mathbf{x}, \mathcal{C})$ directly to find \mathbf{x}^* , one could try to evaluate the mean field $\langle \mathbf{x} \rangle$ at a very low temperature. This is the basic idea behind mean field annealing.

Mean field equations

The exact mean field is computed as

$$\langle \mathbf{x}_i \rangle = \sum_{\mathbf{x}_i \in \mathbb{X}} \mathbf{x}_i P(\mathbf{x}_i) \quad (4.35)$$

but as it is in practice impossible to compute the marginal probabilities $P(\mathbf{x}_i)$, the following approximation [7] is used:

$$\langle \mathbf{x}_i \rangle \approx \sum_{\mathbf{x}_i \in \mathbb{X}} \mathbf{x}_i P(\mathbf{x}_i | \langle \mathbf{x}_{\mathcal{N}_i} \rangle) \quad (4.36)$$

The rationale behind the *mean field approximation* is that the influence of x_j with $j \neq i$ can be approximated by the influence of $\langle x_j \rangle$. Eq. 4.36 can now be rewritten as

$$\langle \mathbf{x}_i \rangle \approx \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} P(x_{i1} = 1 | \langle \mathbf{x}_{\mathcal{N}_i} \rangle) + \dots + \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} P(x_{im} = 1 | \langle \mathbf{x}_{\mathcal{N}_i} \rangle) \quad (4.37)$$

and thus

$$\langle x_{ik} \rangle \approx P(x_{ik} = 1 | \langle \mathbf{x}_{N_i} \rangle) \quad k = 1, \dots, m \quad (4.38)$$

If the conditional probability in Eq. 4.38 is replaced using Eq. 4.31-4.32, a set of nm nonlinear equations in nm unknowns $\langle x_{ik} \rangle$ is obtained:

$$\beta_{ik} = \frac{\exp(-\frac{1}{T}[d_i(\mathbf{f}_{ik}) + \frac{n-1}{t} \sum_{j \in N_i} \sum_{b=1}^m \beta_{jb} d_{ij}(\mathbf{l}_{ik}, \mathbf{l}_{jb})])}{\sum_{a=1}^m \exp[-\frac{1}{T}d_i(\mathbf{f}_{ia}) + \frac{n-1}{t} \sum_{j \in N_i} \sum_{b=1}^m \beta_{jb} d_{ij}(\mathbf{l}_{ia}, \mathbf{l}_{jb})]} \quad (4.39)$$

The unknowns β_{ik} are the estimates of the mean values $\langle x_{ik} \rangle$ and are computed from Eq. 4.39 using iterative resubstitution. In an ideal situation, this set of equations needs to be solved once with a sufficiently small temperature (see Eq. 4.34). However, the convergence domain of the resubstitution algorithm is small for a low temperature. To cope with this, the algorithm is solved several times starting at a sufficiently high T , and gradually decreasing the temperature towards zero.

The MFA algorithm is listed in Algorithm 4.4.3. A set of auxiliary variables α_{ik} is introduced to compute updates for β . During optimization, the means β for which $\sum_{k=1}^m \beta_{ik} = 1$ vary from soft assignments at the higher temperatures to hard assignments (all but one components will be zero) as the temperature reaches zero.

As in each iteration, nm variables α and nm variables β need to be calculated, and because their calculation is linear with m , the computational complexity of one iteration of the resubstitution algorithm is $O(nm^2)$. Unlike DP, the MFA algorithm does not suffer from complex graph configurations.

4.2.3 Heuristic search method

In Section 4.2.1 DP was proposed for optimization. The major advantage of this approach is that it finds a global optimum, but DP is practically only applicable to graphs with limited complexity. Therefore, a third optimization strategy is proposed that is based on DP and where the computation time does not depend on the complexity of the graph.

The basic idea is to look at a set of (overlapping) subgraphs of $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and to optimize these subgraphs individually using DP. Open contours are chosen as subgraphs to keep the computational complexity limited ($O(nm^2)$). Prior to optimization a set of contours (paths) are generated randomly, starting from a specific edge (Fig. 4.3(a)) and iteratively adding concatenating edges till all landmarks are captured (b) or as long as the path is not self-intersecting (c). From every edge in \mathcal{E} a random path is started resulting into a total of t subgraphs $\mathcal{G}_j = (\mathcal{V}_j, \mathcal{E}_j)$. This guarantees that every edge is taken into account for optimization.

Once all the paths are generated, an iterative optimization procedure is started as follows. Each subgraph \mathcal{G}_j , $j = 1, \dots, t$ is optimized globally using Algorithm 4.4.1. The result yields votes for the selected landmark candidates. After optimizing every random path, the least voted candidates are removed

such that the overall number of candidates $n \sum_{i=1}^n m_i$ is reduced with a pre-defined fraction f_{red} . This procedure is repeated till only one candidate is left for every landmark. This optimization scheme is described in Algorithm 4.4.4. As this method iteratively uses DP, it will be referred to as *iterative dynamic programming* (IDP).

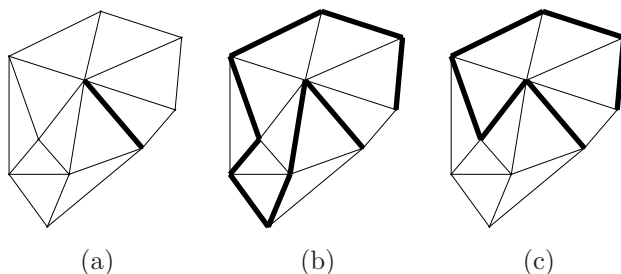


Figure 4.3: A path is generated by randomly adding neighboring points starting from a specific edge (a) till all points are used (b) or till the path becomes selfintersecting (c).

4.3 Implementation issues

Some practical considerations concerning the implementation of the model fitting are discussed here. Section 4.3.1 describes how the discretization is practically done and manages to keep the number of candidates m as small as possible. This is important as for all the proposed optimization algorithms, the computation time increases drastically with m . A second issue, handled in Section 4.3.2, deals with the computation of the intensity energy.

4.3.1 Search regions

As mentioned in Section 4.1.2 the search space is discretized by evaluating the intensity energies $d_i(\mathbf{f}_i)$, $i \in \mathcal{V}$ at a number of locations in the image and selecting the m lowest-energy-candidates for each landmark. This number should be taken large enough to ensure that appropriate locations are captured. When the whole image is scanned for candidates, a larger m is required than when only a relevant part of the image is explored. This is illustrated in Fig. 4.1 and Fig. 4.4. When searching for landmarks of the left lung shape, also candidate locations located at the right lung field will be selected when the whole image is scanned, because similar gray-level patterns can be found there (Fig. 4.1).

To keep the required number of candidate locations as small as possible, the following discretization approach is used. For each landmark, a search region is defined that samples points in that part of the image where the landmark is expected to lie. Points are sampled on a grid with a fixed distance δ between

neighboring points in each direction (x, y and also z for 3D images). The size and the location of these landmark-individual regions of interest (ROI) can be learnt from examples as follows (Fig. 4.4). The ROI is centered around the a priori expected location $\boldsymbol{\mu}_{\mathbf{l}_i} = (\mu_{x_i}, \mu_{y_i}, \mu_{z_i})$ of landmark \mathbf{l}_i and its extent is defined such that each point in the grid $\mathbf{l}_s = (x_s, y_s, z_s)$ deviates at most a fixed number of standard deviations (typically $a = 3$) from the mean:

$$\mu_{x_i} - a\sigma_{x_i} \leq x_s \leq \mu_{x_i} + a\sigma_{x_i} \quad (4.40)$$

$$\mu_{y_i} - a\sigma_{y_i} \leq y_s \leq \mu_{y_i} + a\sigma_{y_i} \quad (4.41)$$

$$\mu_{z_i} - a\sigma_{z_i} \leq z_s \leq \mu_{z_i} + a\sigma_{z_i} \quad (4.42)$$

A second technique to keep the required number of candidates limited, is to apply a multi-resolution search strategy. A first stage starts with sufficiently large search regions centered at the mean positions and sampled with a large grid spacing δ . In later stages, more accurate samplings (smaller δ 's) are used in smaller search regions positioned around the computed locations as found in the previous stages.

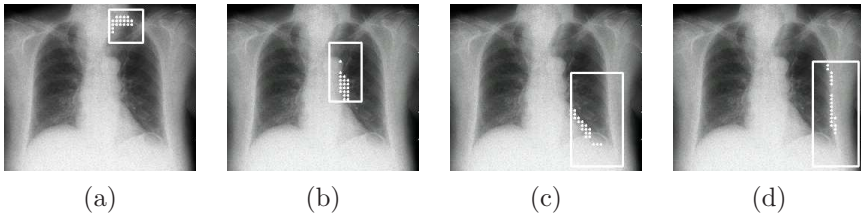


Figure 4.4: Candidate locations for the landmarks shown in Fig. 3.11(a) are searched by evaluating the intensity energies at points of a search grid that overlays that part of the image where the landmark is expected as learnt from the training set. This way only a limited number of candidates m is needed to capture the appropriate locations.

4.3.2 Energy balance

If the intensity model $p(\mathbf{I}|\mathbf{l})$ is implemented as described in Section 3.6.2, a bias towards intensity information is made in the energy function (Eq. 4.10 and Eq. 4.11) due to an overestimation of the intensity energy as illustrated in Fig. 3.11.

The parameter ρ was introduced to bring the intensity energy in a statistically meaningful range. Two approaches are described here to find an appropriate value for ρ .

Using a validation measure

One technique is to segment unseen images for several ρ -values and select the value that leads to the smallest difference between automated outcomes and

ground-truth shapes. It is important to run the segmentation algorithm on unseen images and not on the training images. A first possibility is to apply the algorithm on a second set of images (validation set), for which the ground-truth shapes are known. A second possibility does not need a validation set but uses a leave-one-out strategy on the training images. A model is built from all training images except one and the model is subsequently applied for segmentation on the image that was left out. Doing this for all training images and for several ρ -values enables to find an appropriate value.

Using the energy balance

A second approach considers ρ as part of the statistical model, rather than as a parameter to be tuned. Assume that an image would be segmented with $\rho = 1$ (overestimated intensity energy), the shape energy of a solution that deviates significantly from the shape model would only have a minor contribution in the global object energy. As a result, the optimization algorithm would probably find a shape that has a low intensity energy but not necessarily has a low shape energy.

The statistical distributions $p(\mathbf{l})$ and $p(\mathbf{I}|\mathbf{l})$ enable to compute an expected shape energy $\langle E_{\mathbf{l}} \rangle$ and an expected intensity energy $\langle E_{\mathbf{f}} \rangle$. These energy values are the mean energy values of the training shapes and can be computed directly from the model parameters:

$$\langle E_{\mathbf{l}} \rangle = \frac{n-1}{t} \sum_{\{i,j\} \in \mathcal{E}} \langle d_{ij}(\mathbf{e}_{ij}) \rangle \quad (4.43)$$

$$\langle E_{\mathbf{f}} \rangle = \sum_{i=1}^n \langle d_i(\mathbf{f}_i) \rangle \quad (4.44)$$

It can be shown that the expected value of the Mahalanobis distance between a random vector and its distribution is equal to the rank of the covariance matrix of that distribution (Appendix B). Consequently,

$$\langle E_{\mathbf{l}} \rangle = \frac{n-1}{t} t n_d = (n-1) n_d \quad (4.45)$$

$$\langle E_{\mathbf{f}} \rangle = \frac{1}{\rho} \sum_{i=1}^n \sum_{j=1}^{n_{\mathbf{I}}} \text{rank}(\Sigma_{\mathbf{f}_i^{[j]}}) \quad (4.46)$$

with n_d the number of dimensions of the edge vectors \mathbf{e}_{ij} . Hence, $n_d = 2$ and $n_d = 3$ for respectively 2D and 3D images.

A shape \mathbf{l}^* segmented with a too low value for ρ will on average result in shape energies much higher than $\langle E_{\mathbf{l}} \rangle$ and intensity energies much lower than $\langle E_{\mathbf{f}} \rangle$. A too large ρ would do the opposite. Too much attention would be given to the shape energy. Hence, a shape would be found that lies close to the mean shape, but not properly fits to the image data.

This suggests the following statistical criterion: a proper ρ will result in optimal shapes \mathbf{l}^* that on average bring the shape and intensity energies into the

correct balance:

$$\frac{\langle E_{\mathbf{I}^*} \rangle}{\langle E_{\mathbf{I}} \rangle} \approx \frac{\langle E_{\mathbf{f}^*} \rangle}{\langle E_{\mathbf{f}} \rangle} \quad (4.47)$$

with $\langle E_{\mathbf{I}^*} \rangle$ and $\langle E_{\mathbf{f}^*} \rangle$ the expected shape and intensity energy of an optimized shape \mathbf{I}^* . These energy values depend on the value for ρ and are computed by segmenting each training image (using leave-one-out) and computing the corresponding energies $E_{\mathbf{I}}(\mathbf{I}^*)$ and $E_{\mathbf{f}}(\mathbf{I}, \mathbf{I}^*)$ (Eq. 3.38 and 3.54).

If the left hand side of Eq. 4.47 is larger than the right hand side, ρ should be increased and vice versa. This procedure is repeated until the balance 4.47 is in equilibrium.

4.4 Conclusion

Statistical knowledge about the shape of an object and how this object appears in an image, is used to formulate the segmentation task: find the most likely shape of that object in a given image. The energy function that corresponds to this posterior distribution has to be minimized with respect to the unknown shape. As this objective function has many local optima and a huge amount of degrees of freedom, it is difficult to optimize in the continuous domain with conventional techniques.

To find a solution of the proposed objective function, prior to optimization, the search space is discretized by selecting a number of candidate locations for each landmark individually. The resulting combinatorial optimization problem behaves as an MRF and can be solved using several techniques. A first technique uses dynamic programming to find the global optimum. Unfortunately, the computation time of DP becomes easily unacceptable for graphs with complex configurations. Therefore, two alternative techniques are proposed. Mean field annealing, known in the world of MRFs, tries to minimize the energy of the MRF by estimating the mean field for decreasing Gibbs temperatures. A second alternative technique approaches the problem by iteratively solving minimal cost paths and removing the worst candidates.

As the computational complexity of the presented optimization techniques increases fiercely with the number of candidates selected for each landmark, a search method is proposed that uses landmark-individual search regions and a multi-resolution strategy to keep this number as small as possible.

Algorithm 4.4.1: DP OPEN CONTOUR

1. Eliminate graph

$$Q_0 = 0$$

for $i \leftarrow 1$ to $n - 1$

$$\text{do } \left\{ \begin{array}{l} \text{for each } \mathbf{x}_{i+1} \in \mathbb{X} \\ \text{do } \left\{ \begin{array}{l} Q_i(\mathbf{x}_{i+1}) = \min_{\mathbf{x}_i} [d_i(\mathbf{x}_i) + d_{i,i+1}(\mathbf{x}_i, \mathbf{x}_{i+1}) + Q_{i-1}(\mathbf{x}_i)] \\ R_i(\mathbf{x}_{i+1}) = \arg \min_{\mathbf{x}_i} [d_i(\mathbf{x}_i) + d_{i,i+1}(\mathbf{x}_i, \mathbf{x}_{i+1}) + Q_{i-1}(\mathbf{x}_i)] \end{array} \right. \end{array} \right.$$

2. Trace back

$$\mathbf{x}_n^* = \arg \min_{\mathbf{x}_n} [d_n(\mathbf{x}_n) + Q_{n-1}(\mathbf{x}_n)]$$

for $i \leftarrow n - 1$ to 1

$$\text{do } x_i^* = R_i(\mathbf{x}_{i+1}^*)$$

Algorithm 4.4.2: DP CLOSED CONTOUR

1. Eliminate graph

for each $(\mathbf{x}_2, \mathbf{x}_n) \in \mathbb{X}^2$

$$\text{do } \left\{ \begin{array}{l} Q_1(\mathbf{x}_2, \mathbf{x}_n) = \min_{\mathbf{x}_1} [d_1(\mathbf{x}_1) + \frac{n-1}{n}(d_{12}(\mathbf{x}_1, \mathbf{x}_2) + d_{1n}(\mathbf{x}_1, \mathbf{x}_n))] \\ R_1(\mathbf{x}_2, \mathbf{x}_n) = \arg \min_{\mathbf{x}_1} [d_1(\mathbf{x}_1) + \frac{n-1}{n}(d_{12}(\mathbf{x}_1, \mathbf{x}_2) + d_{1n}(\mathbf{x}_1, \mathbf{x}_n))] \end{array} \right.$$

for $i \leftarrow 2$ to $n - 2$

$$\text{do } \left\{ \begin{array}{l} \text{for each } (\mathbf{x}_{i+1}, \mathbf{x}_n) \in \mathbb{X}^2 \\ \text{do } \left\{ \begin{array}{l} Q_i(\mathbf{x}_{i+1}, \mathbf{x}_n) = \min_{\mathbf{x}_i} [d_i(\mathbf{x}_i) + \\ \quad \frac{n-1}{n}d_{i,i+1}(\mathbf{x}_i, \mathbf{x}_{i+1}) + Q_{i-1}(\mathbf{x}_i, \mathbf{x}_n)] \\ R_i(\mathbf{x}_{i+1}, \mathbf{x}_n) = \arg \min_{\mathbf{x}_i} [d_i(\mathbf{x}_i) + \\ \quad \frac{n-1}{n}d_{i,i+1}(\mathbf{x}_i, \mathbf{x}_{i+1}) + Q_{i-1}(\mathbf{x}_i, \mathbf{x}_n)] \end{array} \right. \end{array} \right.$$

for each $\mathbf{x}_{n-1} \in \mathbb{X}$

$$\text{do } \left\{ \begin{array}{l} Q_{n-1}(\mathbf{x}_n) = \min_{\mathbf{x}_{n-1}} [d_{n-1}(\mathbf{x}_{n-1}) + \\ \quad \frac{n-1}{n}d_{n-1,n}(\mathbf{x}_{n-1}, \mathbf{x}_n) + Q_{n-2}(\mathbf{x}_{n-1}, \mathbf{x}_n)] \\ R_{n-1}(\mathbf{x}_n) = \arg \min_{\mathbf{x}_{n-1}} [d_{n-1}(\mathbf{x}_{n-1}) + \\ \quad \frac{n-1}{n}d_{n-1,n}(\mathbf{x}_{n-1}, \mathbf{x}_n) + Q_{n-2}(\mathbf{x}_{n-1}, \mathbf{x}_n)] \end{array} \right.$$

2. Trace back

$$\mathbf{x}_n^* = \arg \min_{\mathbf{x}_n} [d_n(\mathbf{x}_n) + Q_{n-1}(\mathbf{x}_n)]$$

$$\mathbf{x}_{n-1}^* = R_{n-1}(\mathbf{x}_n^*)$$

for $i \leftarrow n - 2$ to 1

$$\text{do } x_i^* = R_i(\mathbf{x}_{i+1}^*, \mathbf{x}_n^*)$$

 Algorithm 4.4.3: MFA FOR OPTIMAL LABELING

input: $\mathcal{C} = \{\{\mathbf{l}_{ik}, \mathbf{f}_{ik}\}_{k=1}^m\}_{i=1}^n$: candidates T_{start} : start temperature T_{end} : end temperature n_T : number of annealing temperatures**annealing scheme:** $h = (T_{\text{end}}/T_{\text{start}})^{1/(n-1)}$ $T = T_{\text{start}}$ initialize β : $\beta_{ik} = 1/m$ $i \in \mathcal{V}, k \in \{1, \dots, m\}$ **for** $i_T \leftarrow 1$ **to** n_T

$$\text{do } \left\{ \begin{array}{l} \text{while } \beta \text{ not stable} \\ \text{do update } \beta \\ T = hT \end{array} \right.$$
 $\mathbf{x}^* = \beta$ **update** β :**for** $i \leftarrow 1$ **to** n

$$\text{do } \left\{ \begin{array}{l} \text{for } k \leftarrow 1 \text{ to } m \\ \text{do } \alpha_{ik} = \exp\left(-\frac{1}{T}[d_i(\mathbf{f}_{ik}) + \frac{n-1}{t} \sum_{j \in \mathcal{N}_i} \sum_{b=1}^m \beta_{jb} d_{ij}(\mathbf{l}_{ik}, \mathbf{l}_{jb})]\right) \end{array} \right.$$
for $i \leftarrow 1$ **to** n

$$\text{do } \left\{ \begin{array}{l} \text{for } k \leftarrow 1 \text{ to } m \\ \text{do } \beta_{ik} = \frac{\alpha_{ik}}{\sum_{a=1}^m \alpha_{ia}} \end{array} \right.$$

 Algorithm 4.4.4: HEURISTIC SEARCH METHOD

input:

$\mathcal{C} = \{\{\mathbf{l}_{ik}, \mathbf{f}_{ik}\}_{k=1}^m\}_{i=1}^n$: candidates

f_{red} : candidate reduction

generate a number of random paths:

for each $\{i, j\} \in \mathcal{E}$

do generate a path $\mathcal{G}_j = (\mathcal{V}_j, \mathcal{E}_j)$ starting from edge $\{i, j\}$

initial number of candidates:

$m_i = m, \quad \forall i \in \mathcal{V}$

iteratively optimize and prune:

while $\sum_{i=1}^n m_i > n$

do {

initialize votes:

$\mathbf{v}_i = 0 \in \mathbb{R}^{m_i}, \quad \forall i \in \mathcal{V}$

optimize paths:

for $j \leftarrow 1$ **to** t

do {

$\mathbf{x}_{\mathcal{V}_j}^* = \text{DP}(\mathcal{G}_j)$

$\mathbf{v}_i \leftarrow \mathbf{v}_i + \mathbf{x}_i^*, \quad \forall i \in \mathcal{V}$

$\mathbf{v}_i \leftarrow \mathbf{v}_i / \sum_{a=1}^{m_i} v_{ia}$

remove least voted candidates:

$\sum_{i=1}^n m_i \leftarrow (1 - f_{\text{red}}) \sum_{i=1}^n m_i$

 }

Chapter 5

Validation

Together with the model building (Chapter 3) and model fitting (Chapter 4), a number of choices arises. During model construction, a shape representation with landmarks and edges needs to be chosen. Also the construction of the intensity model and the model fitting need decisions to be made and some parameter values to be specified.

This chapter elucidates these choices by making an in-depth validation on lung field segmentation from chest radiographs [8, 11, 36, 38–41, 74]. Each part of the algorithm is analyzed carefully to find appropriate parameter values and to defend the methodological choices. Although these decisions depend on the application, a lot of insight transferrable to other anatomical objects can be gained. The following items are dealt with:

- **Model construction**

- Where to define the landmarks and the edges? (Section 5.2).
- How to determine the parameters of the intensity descriptor (LOIs and profile configuration)? (Section 5.5)
- How to determine a proper ρ -value? (Section 5.4)
- How many training images are needed? (Section 5.6)

- **Model fitting**

- What about the discretization? How is the search grid built? How many candidates have to be selected? (Section 5.5)
- What is the influence of the chosen optimization algorithm? (Section 5.3)
- What is the effect of using local models instead of one global description of the shape? (Section 5.6)

5.1 Lung field data

Two data sets of standard posterior-anterior chest radiographs containing both normal and abnormal cases (Fig. 5.1) were used in this experiment. The first set consists of 50 images and was employed to train and tune the model. The second set contains 44 images and served as a validation set. The width of all images is 256 pixels. The left and right lung field were manually delineated by a trained radiologist in each image of both sets. The manual delineated curves consisted of an arbitrary number of points.

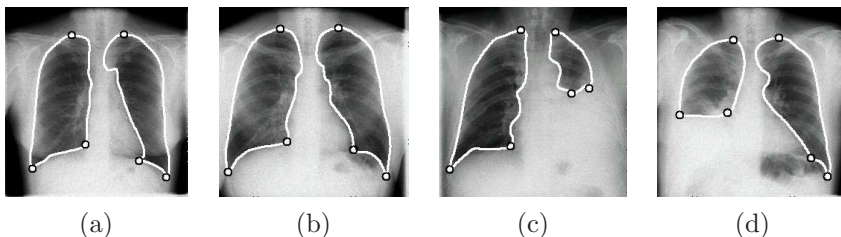


Figure 5.1: The chest radiograph database contains normal (a,b) and abnormal (c,d) cases.

5.2 Object representation

A training set is only complete if landmarks are defined on all shape instances and when the edges \mathcal{E} are established.

Landmarks

Anatomical objects typically have a clearly visible boundary which pleads for points at the object boundary. As the inside of anatomical structures often contains valuable information it might be useful to define additional internal landmarks. However, in this experiment only border-landmarks are considered. Firstly, three easily identifiable landmark points were marked at each lung shape (as indicated in Fig. 5.1). Additional intermediate landmarks were subsequently defined by equidistantly sampling along the contour. Two different representations with either 14 and 40 landmark points were built this way.

Edges

It is a logical choice to place edges between successive landmarks along the contour only. This approach assumes that a point only directly interacts with its first neighbors on the contour:

$$p(\mathbf{l}_i | \mathbf{l}_{\mathcal{V}/\{i\}}) = p(\mathbf{l}_i | \mathbf{l}_{i-1}, \mathbf{l}_{i+1}) \quad (5.1)$$

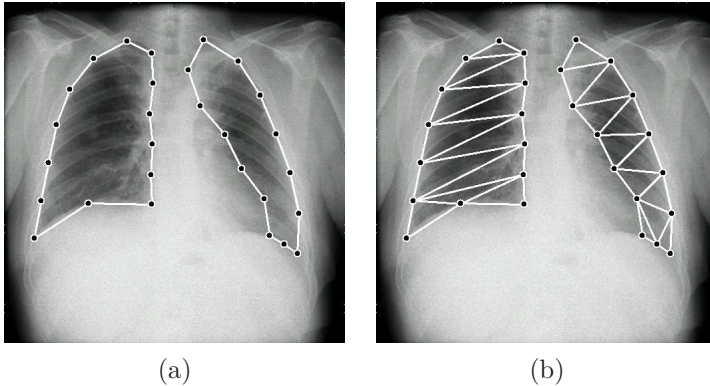


Figure 5.2: Two graph configurations to represent the lung fields. The first one (a) assumes only direct interaction between successive landmarks along the contour. The second graph (b) takes extra dependencies into account that cross the object from one side to the opposite side.

with \mathbf{l}_{i-1} , \mathbf{l}_i and \mathbf{l}_{i+1} three successive points. However, it sounds reasonable that a landmark also interacts with more landmarks than only its first neighbors. Hence, one can wonder whether the assumption (Eq. 5.1) leads to a loss in segmentation performance. This was investigated with the following experiment.

Consider the two configurations shown in Fig. 5.2. The first one (a) assumes only direct interaction between successive landmarks whereas the second graph (b) takes extra dependencies into account that cross the object from one side to the other side. The experiment intends to find out whether the first representation is acceptable and well suited for segmentation. So as to answer this question, a model was built for each graph and applied for segmentation.

In the objective function to be minimized (Eq. 4.10 or Eq. 4.11), the shape potentials that depend on more than two landmarks were approximated as combinations of second order potentials (Section 3.5.3). Consequently, in this experiment the energy of the first configuration is exact, but the energy of the second one is only an approximation. To eliminate the influence of the approximation, this experiment was carried out with exact shape energies (Appendix A) instead and DP was used for optimization.

Models for the two shape configurations were built with the 50 training images using the settings listed in Table 5.1. A small number of pilot experiments was carried out to find an appropriate set of parameter values. The intensity model was constructed using spherical profiles with a radius $r_{\mathbf{f}} = 8$ pixels and $n_{\mathbf{f}} = 4$ samples. A total of $n_{\mathbf{I}} = 60$ feature images were generated with the same LOI parameters as in [8]: the two moments ($M = 1, 2$), a filter bank of all derivatives up to second-order ($L, L_x, L_y, L_{xx}, L_{yy}, L_{xy}$) and five inner scales ($\sigma = 0.5, 1, 2, 4, 8$ pixels) and corresponding outer scales $\alpha = 2\sigma$. The parameter ρ to compensate for the overestimated intensity energy was set to $\rho = 15$.

An elaborate search for an appropriate ρ -value is described in Section 5.4. The following settings were used to fit the model. For each landmark a rectangular search region was centered at the mean landmark location as deduced from the training shapes. The extent of the grid was taken such that the x - and y -coordinates did not deviate more than $a = 3$ times the standard deviation from the mean of x and y respectively. The distance between neighboring search points was set to $\delta = 4$ pixels. In each search region, $m = 50$ candidate locations were selected for discretization.

The quality of an automated segmentation \mathbf{I}^* with respect to a ground-truth segmentation \mathbf{I}^{GT} is computed as the averaged euclidian distance Λ between corresponding landmark points:

$$\Lambda = \frac{1}{n} \sum_{i=1}^n \|\mathbf{l}_i^* - \mathbf{l}_i^{\text{GT}}\| \quad (5.2)$$

A second validation measure computes the overlap between the manual and automated delineations:

$$\Omega = \frac{A_{\text{TP}}}{A_{\text{TP}} + A_{\text{FP}} + A_{\text{FN}}} \quad (5.3)$$

where A_{TP} stands for true positive (area correctly classified as lung), A_{FP} for false positive (area incorrectly classified as lung), and FN for false negative (area incorrectly classified as background). This overlap coefficient always lies between 0 and 1.

The results of the model fitting on the validation images are summarized in Table 5.2. The mean and standard deviation of the overlap coefficient and distance error are given for both configurations. According to these results, it can not be concluded that the incorporation of extra dependencies than only the first neighbors along the contour, improves segmentation. A paired t -test gave no significant difference.

Hence, the choice to define only edges between first neighbors along the border seems acceptable. If however the objects would have random orientations, the assumption (Eq. 5.1) would not make sense. Indeed, the assumption pretends that when both \mathbf{l}_{i-1} as \mathbf{l}_i are known, only \mathbf{l}_i holds useful information for the prediction of \mathbf{l}_{i+1} . It is obvious that when objects are randomly oriented, the location of \mathbf{l}_{i-1} can not be ignored for the prediction of \mathbf{l}_{i+1} .

5.3 Optimization algorithm

In Chapter 4 three optimization techniques were proposed. In this section a number of experiments is worked out as to gain insight in the performance of these algorithms. Dynamic programming always finds a globally optimal solution whereas MFA and IDP do not guarantee a global optimum. One could wonder whether these alternative algorithms in practice really end up in another optimum. And if they do, has this effect on the segmentation performance?

Symbol	Description	Value
Model construction		
n	number of landmarks	2×14 (left and right)
t	number of edges	2×14 or 2×25
$n_{\mathbf{f}}$	profile length	4
$r_{\mathbf{f}}$	profile radius	8 pixels
σ	LOI inner scale	0.5, 1, 2, 4, 8 pixels
α	LOI outer scale	2σ pixels
L_i	LOI derivatives	$L, L_x, L_y, L_{xx}, L_{yy}, L_{xy}$
M	LOI moments	1, 2
ρ	overestimation intensity energy	15
Model fitting		
δ	spacing search grid	4 pixels
a	search grid extent	3
m	number of candidates	50

Table 5.1: Parameters for lung field segmentation.

	Right lungfield		Left lungfield	
	graph 1	graph 2	graph 1	graph 2
Ω (%)	92.4 ± 2.4	92.3 ± 2.5	89.5 ± 6.8	89.2 ± 6.9
Λ (pixels)	8.4 ± 3.3	8.3 ± 3.2	10.3 ± 6.7	9.8 ± 6.9

Table 5.2: The mean and standard deviation of the overlap Ω (in %) and distance error (in pixels) are given for the two configurations of Fig. 5.2. According to these results, it can not be concluded that the incorporation of extra dependencies than only the first neighbors along the contour, improves segmentation.

Experiments

The three algorithms were applied for different graph configurations. As the DP algorithm easily gets computationally too involved, configurations with a limited complexity (maximally $O(nm^3)$ with small n and limited m) were chosen. As a result, DP could be applied for all configurations and consequently, the ability of MFA and IDP to end up in the global optimum could be verified. The experiment was done on the three configurations shown in Fig. 5.3. The first and second one represent the lung fields as respectively open and closed contours. The third one is a triangulated graph that incorporates extra edges similar as in Section 5.2.

Models were constructed for each shape configuration using the settings of Table 5.1. The three models were fitted to the validation images using dynamic programming (DP), mean field annealing (MFA) and iterative DP (IDP). In Section 4.2.1, an alternative DP algorithm to optimize closed contours was

proposed that runs in $O(nm^2)$ instead of $O(nm^3)$ but doesn't guarantee a global optimum (DPCC). The performance of this algorithm on the closed contour representation will be reported also. The required parameters for the model fitting are given in Table 5.3. A small number of experiments was carried out to find appropriate values for both MFA and IDP. A proper range of annealing temperatures was established as follows. During annealing, the mean field β evolves with a changing Gibbs temperature T . It seemed that when the start temperature was set too high, β did not advance at all during the first iterations. Also when the temperature becomes too small, the mean field reaches a steady state. A proper range seemed $T \in [10^{-3}, 10^6]$.

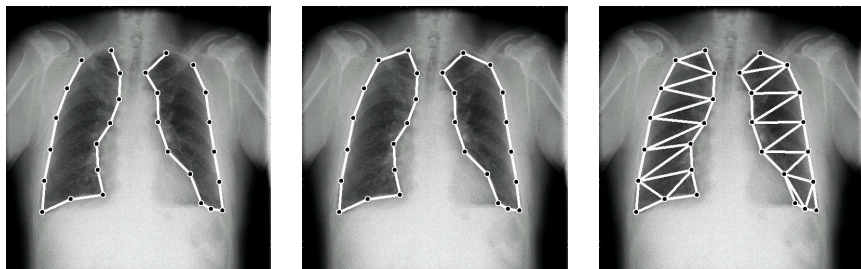


Figure 5.3: Three different graph configurations are tested. The first and second one, represent the lung fields as respectively open and closed contours. The third one is a triangulated graph that incorporates connections that do not coincide with the lung boundaries.

Symbol	Description	Value
Optimization		
T_{start}	MFA start temperature	10^6
T_{end}	MFA end temperature	10^{-3}
n_T	MFA number of annealing temperatures	100
f_{red}	IDP candidate reduction fraction	0.05

Table 5.3: Parameters for lung field segmentation.

Results

The optimal energies found with DP, IDP, MFA and DPCC averaged over the 44 validation images are listed in Table 5.4. The difference between on one hand the optimal energies found with MFA, IDP and DPCC and on the other hand the global minimal energy (found with DP) is shown in Fig. 5.4 for (a) open contour, (b) closed contour and (c) triangulation. The plotted values ϵ

are relative energy differences:

$$\epsilon = \frac{E(\mathbf{x}) - E(\mathbf{x}^*)}{E(\mathbf{x}^*)} \quad (5.4)$$

where \mathbf{x} is the solution found with MFA or IDP. DP finds the global minimum for the three configurations, IDP also finds the minimum of the energy function for the open contour. For the closed contour, MFA, IDP and DPCC found the global optimum for respectively 23, 27, and 22 out of 44 images. However, if we take a look at Fig. 5.4(b) it seems that the DP based methods score significantly better than MFA.

The opposite is observed for the triangulation. MFA found for 24 images the global solution and always ended up in a lower energy as compared to IDP (Fig. 5.4(c)). This experiment suggests that MFA optimizes more easily as landmarks have more neighbors. This is illustrated with the example segmentation shown in Fig. 5.5. The left lung field does not get into the global optimum if it is represented as a contour, but it correctly converges to the true optimum for the triangulated graph.

	DP	MFA	IDP	DPCC
Left lungfield				
open	181.1 ± 39.3	183.9 ± 40.8	181.1 ± 39.3	
closed	181.2 ± 39.1	183.8 ± 40.5	181.2 ± 39.1	181.6 ± 39.4
tri	182.9 ± 37.5	183.8 ± 38.4	187.6 ± 41.0	
Right lungfield				
open	194.3 ± 37.8	194.5 ± 37.9	194.3 ± 37.8	
closed	194.9 ± 37.6	195.1 ± 37.8	194.9 ± 37.6	195.1 ± 37.6
tri	195.1 ± 39.0	195.1 ± 39.0	202.1 ± 44.4	

Table 5.4: The optimal energy of the left and right lung fields found with DP, IDP, MFA and DPCC averaged over the 44 validation images. DP guarantees the global optimum, but this is not true for MFA, IDP and DPCC. For the contours, the DP based methods do better than MFA. The opposite is noticed for the triangulated graph.

A whisker plot of the distance errors for each segmentation scheme (three algorithms and three configurations) is depicted in Fig. 5.6. The boxes indicate the value of the median and the lower and upper quartiles. The vertical line shows the extent of the errors and outlier errors are indicated as red crosses. A paired t -test gave no significant differences ($p < 0.05$) among the three different algorithms and also between the different graph configurations.

In Fig. 5.7 the automated delineations using DP, MFA and IDP on four validation images are shown. The images were selected by first ranking them according to the optimal energy (using DP), and taking the lowest and highest energies (#1 and #44) and two intermediate cases (#15 and #29). A

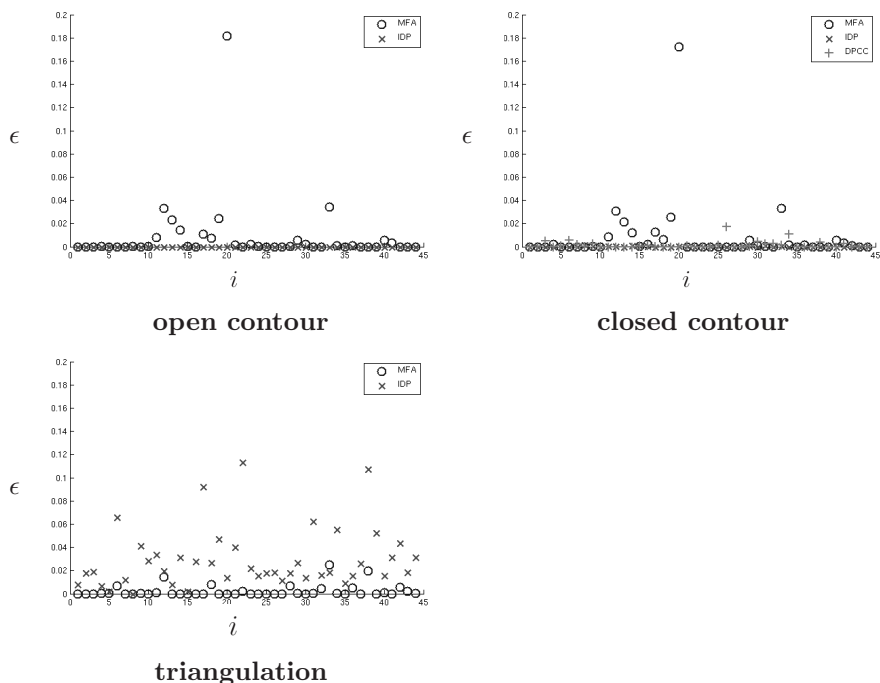


Figure 5.4: The relative deviation ϵ of the energies found with the approximating methods from the globally minimal energy. These plots show that the IDP and DPCC are better suited for contour segmentation, whereas for the triangulated graph, MFA finds segmentations with lower energies than IDP.

segmentation that results in a high energy often corresponds to a pathological case or a wrong segmentation. The third patient has an unusually large heart and for the highest energy case, the left lung is highly pathologic. One landmark of the right lung in the same example has been located wrongly. This fault was due to the large radius used for the intensity profiles. If the profile is positioned at the correct location, the circle gets out of the image range. Using a more accurate model (with smaller profiles) avoided this problem.

5.4 Energy balance

In Section 4.3.2 two techniques were described to find a ρ -value that brings the intensity energy to the right scale. First, the validity of the energy balance criterion will be verified. Subsequently, an appropriate value will be searched using both techniques.

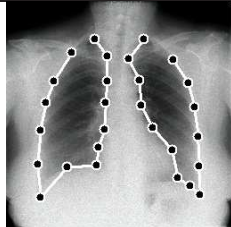
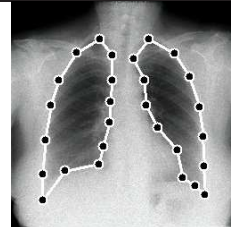
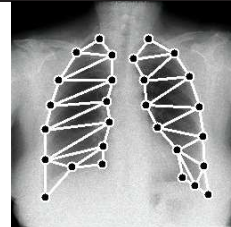
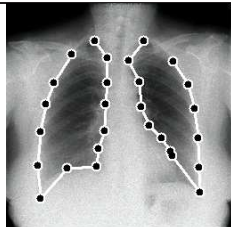
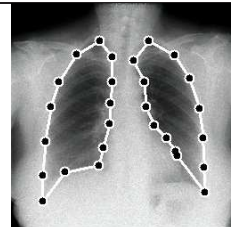
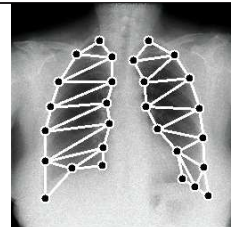
	open contour	closed contour	triangulation
DP	 $E = 296.7$	 $E = 297.9$	 $E = 295.3$
MFA	 $E = 350.5$	 $E = 349.1$	 $E = 295.3$

Figure 5.5: Segmentation of an example image with DP and MFA for the three configurations. MFA does not find the global optimum for the left lung field if it is represented as a contour, but it correctly converges to the true optimum for the triangulated graph.

Experiment 1

A first experiment was set up to test the validity of the proposed energy balance (Eq. 4.47). The same graph configurations as in the previous experiment (Section 5.3) were used. As they all consist of the same 14 landmark points, the same ρ value should be found for each of the three representations.

The criterion was verified as follows. An intensity model was constructed with the same parameters as the previous experiment (Table 5.1) except that only one single LOI is used: one zero-order derivative with $\sigma = 1$, $\alpha = 2$, and only the first statistical moment. It is trivial that for this model the assumption of Eq. 3.52 is valid and consequently it should hold that $\rho = 1$. This theoretical value will be compared with the experimental value that follows from the energy balance criterion described in Section 4.3.2.

The following iterative procedure was followed to bring the balance into equilibrium. A particular ρ is updated according to

$$K_\rho = \frac{\langle E_{1^*} \rangle / \langle E_1 \rangle}{\langle E_{f^*} \rangle / \langle E_f \rangle} = \frac{\langle E_{1^*} \rangle \langle E_f \rangle}{\langle E_{f^*} \rangle \langle E_1 \rangle} \approx 1 \quad (5.5)$$

The expected shape and intensity energies (Eq. 4.45 and 4.46) are

$$\langle E_1 \rangle = 2(n - 1) = 26 \quad (5.6)$$

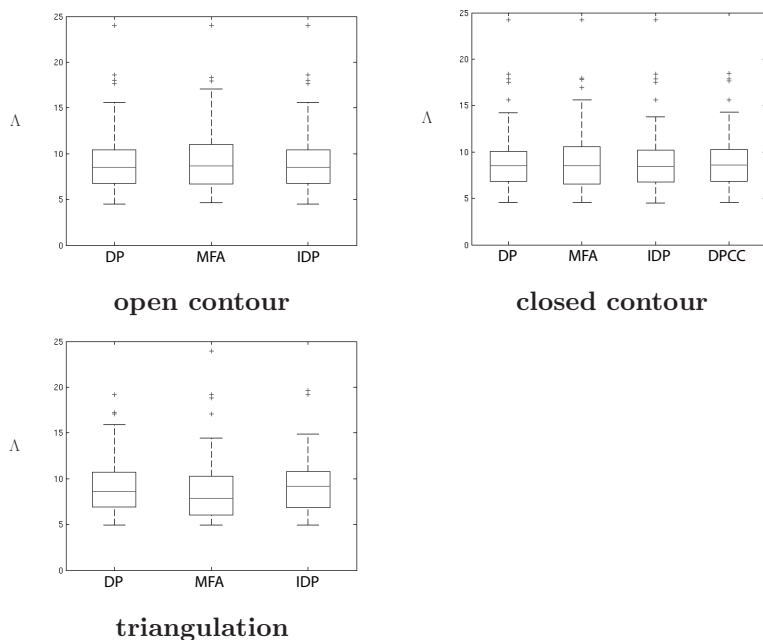


Figure 5.6: Whisker plot of the distance error Λ (pixels) obtained with DP, MFA and IDP for three graph configurations. The boxes indicate the value of the median and the lower and upper quartiles. The vertical line shows the extent of the errors. The crosses indicate outlier errors. No significant differences in segmentation accuracy among the three different optimization algorithms can be observed.

$$\langle E_{\mathbf{f}} \rangle = \frac{1}{\rho} \sum_{i=1}^n \text{rank}(\Sigma_{\mathbf{f}_i^{[1]}}) = \frac{n(n_{\mathbf{f}} - 1)}{\rho} = \frac{42}{\rho} \quad (5.7)$$

The rank of the covariance is $n_{\mathbf{f}} - 1$ because of the profile normalization: the values in the profile are rescaled such that the sum of absolute values equals one. As a result, the profiles have only three degrees of freedom instead of four.

The mean segmentation energies $\langle E_{1^*} \rangle$ and $\langle E_{\mathbf{f}^*} \rangle$ depend on ρ and are obtained by running the segmentation algorithm on the training images using leave-one-out and evaluating the energy functions Eq. 3.38 and Eq. 3.54.

A leave-one-out experiment on the training images that delivers a K_{ρ} considerably higher than 1 means that the energy function gives too much weight to

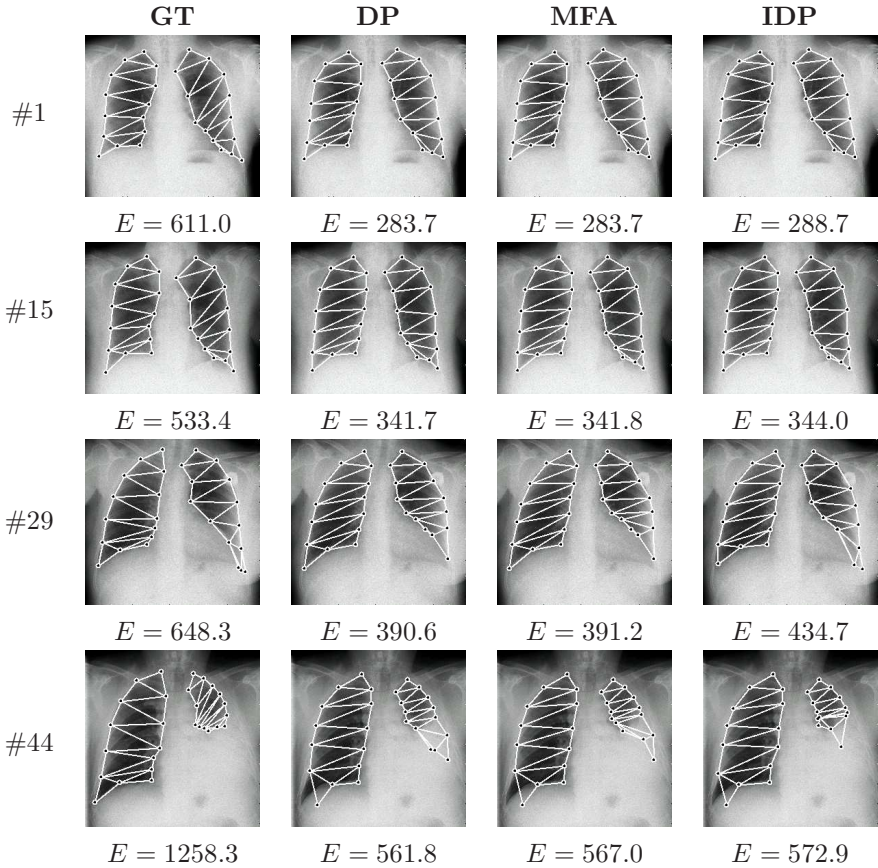


Figure 5.7: Automated delineation using DP, MFA and IDP of images 1, 15, 29 and 44 ranked from low to high energy. The high energies of the third and last image can be attributed to pathologies. The heart in image #29 is unusually large and for the last segmentation, the left lung is highly pathologic. The reported energies are the total object energies (Eq. 4.3) of left and right lung fields together. For comparison, the expected energy for both lungs together is $\langle E \rangle = 388$.

the intensity energy and hence, ρ should be increased:

$$\rho \leftarrow \rho + \Delta_\rho \quad (5.8)$$

with $\Delta_\rho > 0$ a predefined step. If at the other hand $K_\rho < 1$, ρ should be decreased instead. If the direction of the step is changed between two iterations, Δ_ρ is divided by two. This procedure was started from $\rho = 0.1$ and $\Delta_\rho = 5$. The procedure can be initialized from any start point, but this value was chosen such that ρ -values at both sides of the optimal value would be tested. The procedure was interrupted if the difference between two successive

ρ 's became smaller than 0.01.

Result 1

The values found for ρ are reported in Table 5.5. If $m = 50$ candidates were selected for each landmark, the search ended up in ρ -values considerably higher than the expected value $\rho = 1$. This is explained by the fact that the model implicitly assumes that the correct solution is included in the candidate set. For this model, only one feature image was used which made the intensity descriptor less powerful and consequently a higher m was required to ensure that for each landmark proper candidates are selected. Consequently, a shape segmented with $m = 50$ has on average a shape energy that is higher than normal because the optimization algorithm was not able to select candidates that fit the shape model properly. As a result, the ρ -search algorithm brings the balance into equilibrium by increasing ρ . The large differences between left and right lung for $m = 50$ (for example $\rho = 6.62$ versus $\rho = 1.31$) can be explained by the fact that in general the right lung is much more easy to segment. In other words: selecting only 50 landmarks is almost enough for the right lung whether this is not true for the more difficult left lung.

It was found that with $m = 200$, for each landmark the correct landmark locations were selected. Repeating the experiment with $m = 200$ yielded ρ -values close to the theoretical value. Fig. 5.8 shows the value of K_ρ as a function of ρ for the open contour (blue), closed contour (red) and triangulation (green) for both left (a) and right lung field (b).

	$m = 200$		$m = 50$	
	left	right	left	right
open	1.00	0.92	6.62	1.31
closed	1.15	1.00	8.38	1.47
tri	1.08	1.04	6.47	1.47

Table 5.5: The ρ -values that bring the energy balance into equilibrium for the segmentation of left and right lung fields represented as contours and triangulated graphs. If enough candidates are selected ($m = 200$), the experimentally found values agree with the theoretical value $\rho \approx 1$.

Experiment 2

In a second experiment, the energy balance criterion was applied to find the unknown ρ for the model as specified in Table 5.1. As for this model $n_{\mathbf{I}} = 60$, a larger $\rho \in [1, 60]$ is expected. The algorithm was initialized with $\rho = 1$ and $\Delta_\rho = 10$.

A ρ -value found as such is the best choice from a statistical point of view, but one can wonder whether it is the best choice with respect to segmentation performance. To analyze this, the distance error (Eq. 5.2) and overlap coefficient

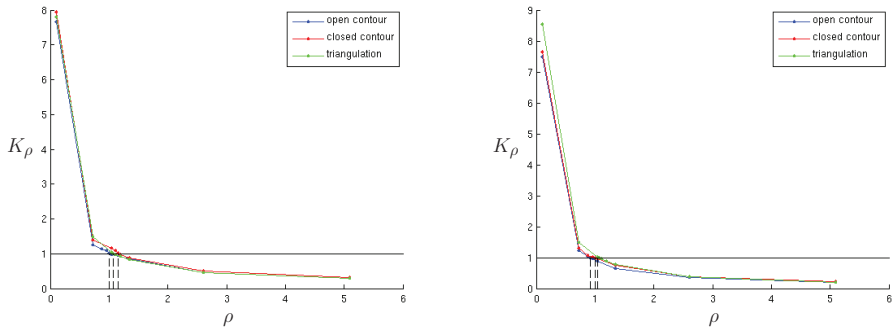


Figure 5.8: K_ρ as a function of ρ with $m = 200$. As expected the curves approximately cross the point $(1, 1)$.

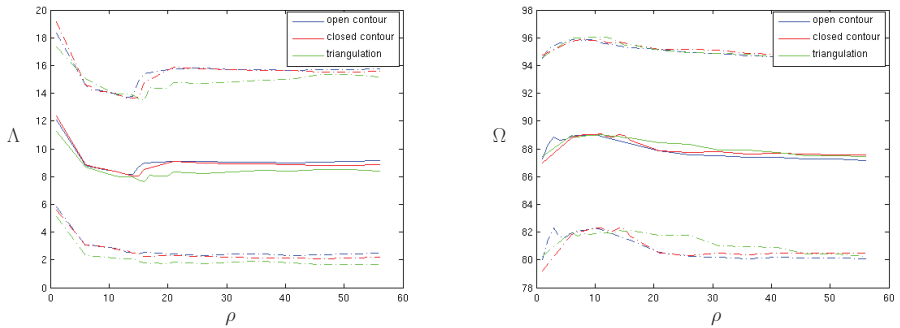


Figure 5.9: Average distance error Λ (pixels) and overlap Ω (%) using leave-one-out on the training set as a function of ρ . The standard deviations are indicated by the dotted lines.

(Eq. 5.3) were computed for several ρ -values.

Result 2

The obtained ρ -values were $\rho_{\text{open}} = 16$, $\rho_{\text{closed}} = 17$ and $\rho_{\text{tri}} = 20$ for the left lung represented as respectively an open contour, a closed contour and a triangulation. For the right lung the values were more consistent: $\rho_{\text{open}} = 15.1$, $\rho_{\text{closed}} = 15.8$ and $\rho_{\text{tri}} = 15.4$.

Fig. 5.9 shows the average distance error Λ (a) and overlap Ω (b) as a function of ρ . It seemed that using a validation criterion to determine ρ has some drawbacks. First, the result might depend on the validation criterion used. Secondly, the curve that shows the average error versus ρ has often meaningless local peaks that make the search for an appropriate value badly conditioned.

5.5 Intensity descriptor

Until now no attention was spent to the parameters of the intensity model. However, several choices have to be made when implementing the descriptor $\mathbf{f}_i = F(\mathbf{I}, \mathbf{l}_i)$. The one presented in Section 3.6.2 needs a set of LOI parameters and a profile configuration ($r_{\mathbf{f}}$ and $n_{\mathbf{f}}$).

5.5.1 LOI parameters

The idea behind the selection of LOI parameters is the following. If a large set of LOIs is computed, probably only a limited set of them will significantly contribute in the detection of landmark candidates, because some features are not able to characterize a certain gray-level pattern. For example, if the image pattern around a landmark point is characterized by an edge in the image along the y -axis, the operator L_x will extract much more relevant information than L_y would do. This suggests that for each individual landmark a set of optimal features needs to be selected. For the ASMOF algorithm, van Ginneken *et al.* [8] used a sequential feature forward selection procedure to find at most 10 LOIs for each landmark, followed by a sequential feature backward selection to prune the set: features are removed if that improves the segmentation performance. This way they could improve the performance of their kNN classifier.

For the proposed image appearance model, in contrast with the ASMOF approach, no segmentation performance can be gained by reducing the LOI set to only a limited set of relevant features. To understand this, consider again how the intensity energy is computed (Eq. 3.53). The contribution of a particular LOI in the total intensity energy is computed as the Mahalanobis distance between the observed profile and the expected profile for that LOI.

Assume a specific LOI that doesn't contain valuable information for a particular landmark. The profiles sampled from this LOI will not be consistent and specific among the training examples and as a result the corresponding PDF will not be sharp (specific). Consequently, every possible profile in a test image will get a low energy value. Such an LOI will not be able to discriminate between a proper and a wrong candidate location, but it also won't interfere with the energy obtained from other LOIs.

This is illustrated in Fig. 5.10 for a landmark (a) along the boundary of the left lung. Two LOIs are considered here. The first one (b) is the first moment of L_x ($\sigma = 1$, $\alpha = 2$) and attenuates edges in the image that run along the y -axis. The second one (c) is the L_y equivalent. The intensity energies of L_x and L_y , computed along a horizontal line that crosses the true landmark location, are shown in respectively Fig. 5.10(e) and Fig. 5.10(f). As expected, the energy of L_x has a strong discriminative power compared to L_y . The total energy is plotted in Fig. 5.10(d) and shows that the low energy values of L_y do not weaken the positive effect of L_x .

As the selection of only the relevant features does not improve segmentation accuracy and also because the selection of an optimal feature set for every

landmark requires an exhaustive procedure, we prefer to use an extensive LOI set for segmentation.

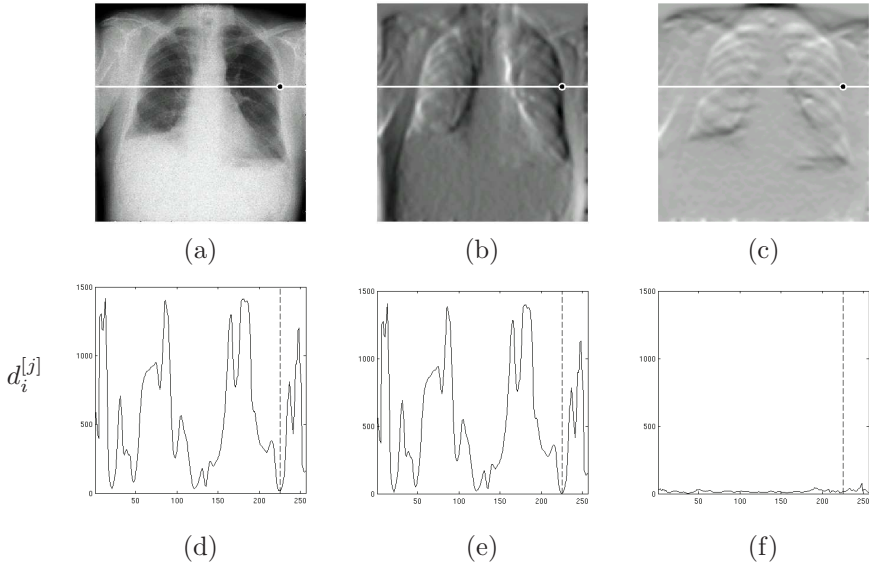


Figure 5.10: The intensity energies corresponding to two different LOIs are computed along a horizontal line that crosses the true landmark location shown in (a). The first LOI (b) is the first moment of L_x ($\sigma = 1, \alpha = 2$) and characterizes edges in the image that run along the y -axis. The second one (c) is the L_y equivalent. Obviously, the energy of L_x (e) has a strong discriminative power compared to L_y (f). The total energy is shown in (d) and proves that the low energy values of L_y do not weaken the positive effect of L_x .

5.5.2 Profile length

The configuration of the spherical profile is determined by two parameters: n_f (profile length) and r_f (profile radius). First, the influence of the number of samples n_f is investigated. Fig. 5.11 shows the tested configurations. For each of them, an individual ρ had to be determined: the higher n_f , the more information is extracted from one LOI and the higher the statistical dependency between different LOI profiles might be. This is confirmed by Table 5.6, that shows ρ found with the energy criterion as a function of n_f .

The segmentation quality of each configuration was tested by running the algorithm on the 44 validation images and computing the overlap and distance error (Fig. 5.12) between manual and automated outcomes. The smaller profiles are preferable as they have a lower computational complexity (Eq. 3.51) and because they do not perform significantly worse than the larger profiles do.

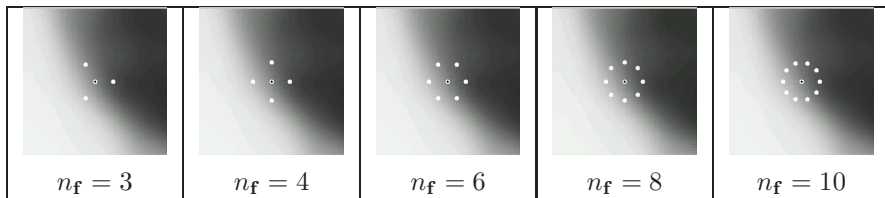


Figure 5.11: Profile configurations with varying length to be compared for lung field segmentation.

n_f	3	4	6	8	10
ρ	11.4	15.3	21.9	26.9	30.6

Table 5.6: The more samples a profile contains, the higher the dependency between different LOIs and thus the more the intensity energy is overestimated.

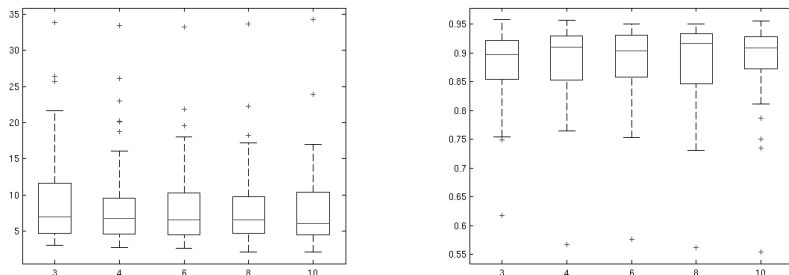


Figure 5.12: Distance error Λ (pixels) and overlap Ω for the 50 training images using leave-one-out for several profiles configurations: $n_f = 3, 4, 6, 8, 10$.

5.5.3 Profile radius

First, a small number of pilot experiments was carried out to find a range of more or less appropriate values: $r_f \in [1, 20]$ pixels. Going outside this range resulted in completely wrong solutions. The choice for a radius (in this range) was important with respect to accuracy and robustness. Accuracy is related to how close the automated solution comes to the ground-truth shape and robustness refers to how likely the algorithm generates completely wrong segmentations.

The relationship between profile radius, accuracy and robustness can be understood as follows. An intensity model built with a larger profile (large r_f) includes information about a larger part of the image around the landmark whereas a smaller profile looks more locally at the image data. Consequently, the smaller ones are able to find the landmark positions more accurately but at the other hand, as they only consider a small neighborhood, they are more likely to select completely wrong candidate locations (false positives). The

larger ones, at the other hand, are more robust but less accurate. This is illustrated with the following experiment.

First, the intensity energy $d_i(\mathbf{I}, \mathbf{l})$ (with a particular r_f) was computed for every pixel in the image. Secondly, the image pixels were sorted according to the energy from low to high. The quality of the intensity descriptor was measured by looking at the ranking of the true landmark location \mathbf{l}_i . Repeating this for every training image (after leaving it out of the model), and for every landmark enabled to quantify the performance of a particular r_f .

Again the same LOI parameters as listed in 5.1 were used. The profiles consisted of $n_f = 4$ points sampled on a circle with a size r_f ranging from 1 to 20 pixels.

Fig. 5.13 shows the relative ranking (in %) averaged over all images and landmarks. Obviously, the larger profiles seem to score significantly better (lower ranking) than the small ones. Fig. 5.14 shows the intensity energy of a particular landmark computed at every pixel location for (a) $r_f = 4$, (b) $r_f = 12$ (c) and $r_f = 18$ (d) pixels. Apparently, the smaller profiles select locations at several regions in the image whereas the larger ones, mostly generate locations in the neighborhood of the true location.

However, if the lung fields were segmented with the same r_f -values, it seemed that the smaller ones performed significantly better than the larger ones. This was experienced after performing a leave-one-out experiment on the training set with $\delta = 2$ pixels, $m = 100$ and $n_f = 4$ and the same LOIs as in the previous experiments. Fig. 5.15 shows the overlap score for different radii and indicates that despite the fact that the smaller profiles result in less robust candidate selectors, their segmented shapes were both accurate and robust. The use of landmark-individual search regions and the incorporated shape knowledge into the energy function, avoid the algorithm to select wrong candidates. This is of course only true if enough candidate locations were selected for each landmark. Hence, the optimal segmentation scheme is one with a sufficiently small grid-spacing ($\delta = 1$ pixel), an accurate profile (for example $r_f = 4$ pixels) and a sufficiently high m . However, such a scheme is not very efficient from a computational point of view. For the candidate selection step, the number of evaluations of the intensity energy function $d_i(F(\mathbf{I}, \mathbf{l}))$ is proportional to $1/\delta^2$ and the optimization step is proportional to m^2 as the DPCC method was applied here.

Instead of using only one accurate resolution level, the same accuracy but with less computation time can be obtained using the multi-resolution strategy as specified in Table 5.7. The algorithm is started with $\delta = 4$ and $r_f = 8$ pixels and sufficiently large search regions. In the following resolution steps, smaller search regions (width and height are divided by two) are centered around the estimated locations from the previous step and are sampled with a smaller spacing and more accurate profiles. This way, the number of evaluations of the intensity energy and the number of selected locations m is reduced without losing quality.

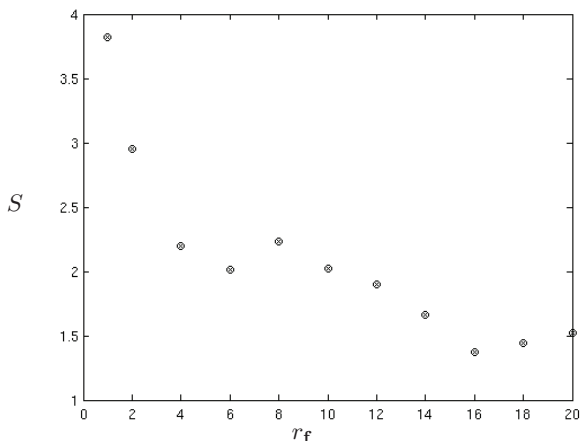


Figure 5.13: Comparison of different profile radii (pixels) for the detection of landmark points. The scores S reported here are obtained as follows. First, the intensity energy for a particular landmark is computed at every pixel location. Secondly, the pixels are ranked according to their intensity energy. The values shown here are the relative ranking (in %) of the ground-truth location.

resolution	δ (pixels)	r_f (pixels)
1	4	8
2	2	4
1	1	4

Table 5.7: Multi-resolution parameters for lung field segmentation.

5.6 Global versus local

In Chapters 1 and 2 the choice for local shape models was motivated. It was claimed that their higher flexibility as compared to the global models enables a better fit to unseen objects. So as to investigate this assumption, a comparison between the proposed segmentation scheme and ASM segmentation is made here.

Experiments

van Ginneken [8] *et al.* applied an ASM scheme for lung field segmentation by using an augmented intensity model based on LOIs and a kNN classifier. In this experiment, the proposed method is compared to this algorithm (AS-MOF) and to a third ASM scheme that was obtained by embedding the same intensity model as the MISC method into the conventional ASM scheme. One change was made to this model: the circular profiles were replaced by

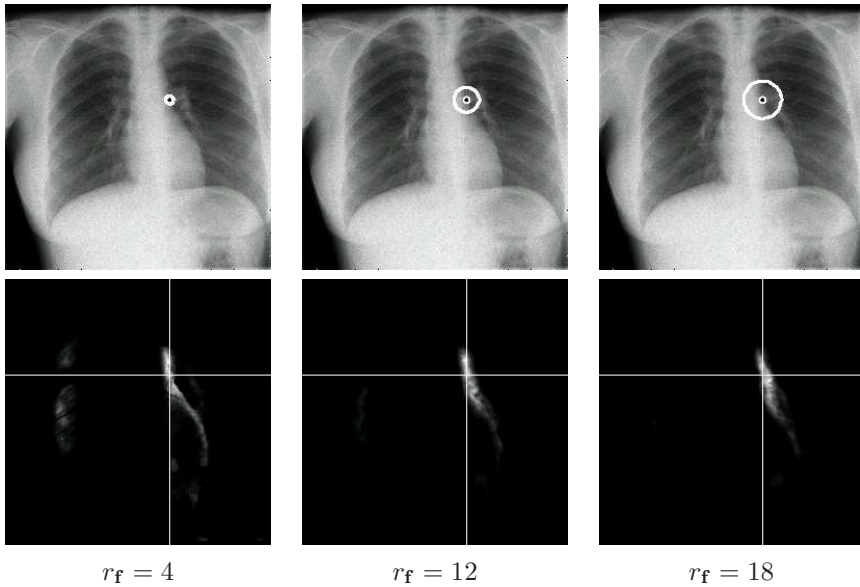


Figure 5.14: Intensity energy of a particular landmark computed at every pixel location for (a) $r_f = 4$, (b) $r_f = 12$ (c) and $r_f = 18$ (d) pixels. Obviously, the smaller profiles would select locations at several regions of the image whereas the largest one only has low energy pixels in the neighborhood of the true location.

linear profiles perpendicular to the contour with k points at either side of the landmark as in [55]. During search, landmark points are updated by evaluating the intensity cost at $2n_s + 1$ locations and selecting the candidate with the lowest intensity cost. This algorithm will be denoted with ASM^* . Both ASM methods were initialized with the mean shape.

For $MISCG$, the settings as obtained from the previous experiments were used. Both the left and right lung fields were represented as closed contours, consisting of $n = 14$ and $n = 40$ landmarks. The intensity models were built with the conventional LOI parameters, spherical profiles with $n_f = 4$. Three resolution levels for segmentation were carried out as specified in Table 5.7.

For $ASMOF$, the same settings as in [8] were used. A number of experiments was performed to find the best performing settings for ASM^* . It was found that a profile with $k = 4$ points at each side of the landmark yielded the best results. Three resolution levels were used and $n_s = 5$ candidate locations at each side of the current location were evaluated.

Three validation measures were used to compare automated and ground-truth segmentations. The overlap Ω and distance error Λ were previously explained. A third validation criterium Δ measures the distance between the two curves. For each point on the first contour, the closest distance to the second contour is computed. As to obtain a one-dimensional measure, the distance error along

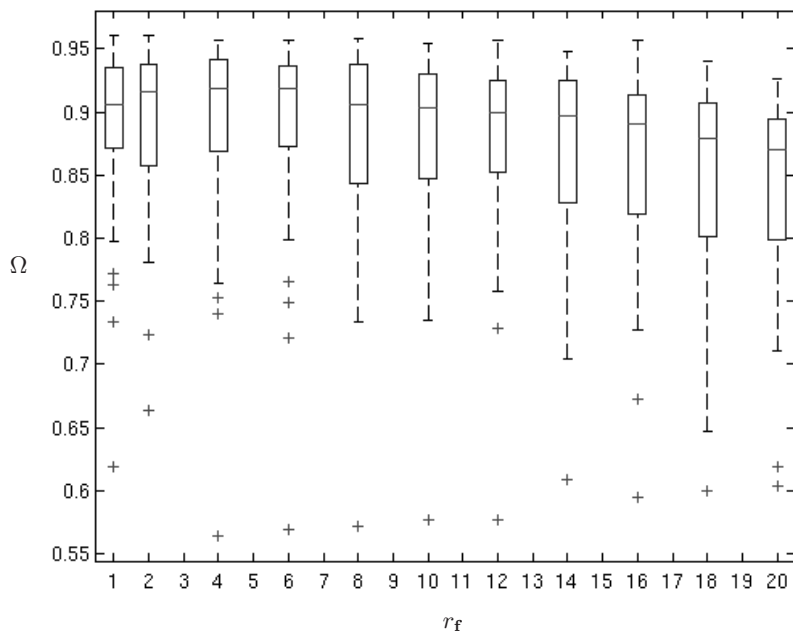


Figure 5.15: Overlap between ground-truth contours and automated segmentations for several r_f -values. The smaller profiles perform significantly better than the larger ones as they allow to detect the lung boundaries more accurately.

the contour is averaged. The error is calculated a second time by switching the role of first and second contour. The mean of the two errors delivers a final symmetrical measurement.

Also the influence of the number of training examples on the segmentation performance was studied. For both the MISCG and ASM* algorithm, the overlap measure was computed for the segmentation of the right lung field, applying models trained with 5 to 50 images.

Results

The results of applying the MISCG, the ASMOF and the ASM* algorithm to the 44 validation images are summarized in Table 5.8. With 14 landmarks, the average overlap for the left and right lung is 90.1% and 92.5% respectively for MISCG, compared to only 84.1% and 86.8% respectively for the ASMOF algorithm. The ASM* method also scores worse than MISCG with 85.6% and 89.5%. The results increase to 92.2% and 93.9% for the MISCG method, to 87.1% and 89.4% for ASMOF and to 88.6% and 92.1% for ASM* when 40 landmarks are used. MISCG also generates the lowest average distance er-

rors, followed by ASM* and ASMOF. For all the results of Table 5.8, MISCG scored significantly better than ASMOF and ASM* with $p < 0.01$ according to a paired t -test. Fig. 5.16 shows the local variation of the distance error along the lung field contours generated by the MISCG, ASMOF and ASM* algorithm with $n = 14$ and $n = 40$ landmarks. Apparently, the ASM algorithms are less accurate at the lower regions (at the mediastinum) than MISCG. Fig. 5.17 shows the results of the three algorithms on five validation images ranked from easy to difficult. The difficulty ranking was achieved as follows. For each image the overlap measure of the three algorithms with respect to the manual segmentations was averaged. The images were ranked according to the mean overlap from high to low. The images at the positions 1, 11, 22, 33 and 44 were selected.

At this point, one can wonder why MISCG performs better than ASM for lung field segmentation. Is it the fact that ASM gets stuck in local optima instead of finding the global optimum? Or is the global shape model too stiff to fit properly to the unseen image data? The following experiment was set up to investigate this (Table 5.9). First, to see whether ASM got stuck in a local minimum, we tried to find ASM solutions that had both a higher performance and a lower value of the intensity cost function. This was achieved by initializing ASM* with the ground-truth (GT) contour. The performance of ASM*GT (91.2%) was significantly better than ASM* (89.8%) and lower intensity costs¹ were achieved (1.07 vs. 1.17). Hence, finding the global optimum would improve the performance, but is still insufficient to explain the difference between MISCG and ASM* (93.0% vs. 89.8%). There seems to be a second problem. ASM tries to find the shape in the PCA subspace with the lowest intensity cost $E_f(\mathbf{1})$. Fitting the GT curve into the PCA space, yields the contour with the highest performance (93.5%) that satisfies the ASM shape constraint (GTPCA in Table 5.9). Unfortunately, the intensity costs of these contours are far from optimal (compare with the intensity costs achieved by ASM*) and this makes that ASM* shifts away from these contours to decrease the intensity cost (from 1.21 to 1.07) and unintentionally lowers segmentation accuracy (from 93.5% to 91.2%). Consequently, the fact that the PCA space is too narrow explains another part of the lower ASM performance. A possible solution would be to increase the flexibility of the shape model by increasing the number of training images.

Fig. 5.18 illustrates the influence of the number of training images on the segmentation performance. Overlap measures for the right lung with $n = 14$ (a) and $n = 40$ (b) landmarks are reported. The required number of training images can be obtained using a paired t -test as follows. One starts from the maximum number i.e. $s = 50$ and gradually decreases this number as long as this does not result in a significant worse result. For the $n = 40$ landmarks and at the 5% level, MISCG needed $s = 10$ images whereas ASM* needed $s = 25$ images. These numbers increased to $s = 20$ and $s = 35$ images respectively at the 15% level. For $n = 14$ landmarks, no difference between MISCG and

¹The reported energy values are normalized such that the expected value equals one.

ASM* could be observed.

(a)

Algorithm	Left lung	Right lung
$n = 14$ landmarks		
MISCG	89.8 ± 6.0	93.0 ± 3.3
ASMOF	84.1 ± 9.0	86.8 ± 8.6
ASM*	85.6 ± 7.9	89.5 ± 4.5
$n = 40$ landmarks		
MISCG	92.2 ± 4.3	94.1 ± 3.1
ASMOF	87.1 ± 9.8	89.4 ± 9.4
ASM*	88.6 ± 7.5	92.1 ± 3.3

(b)

Algorithm	Left lung	Right lung
$n = 14$ landmarks		
MISCG	2.47 ± 1.78	1.82 ± 0.78
ASMOF	4.04 ± 3.00	3.48 ± 2.44
ASM*	3.63 ± 2.36	2.86 ± 1.39
$n = 40$ landmarks		
MISCG	1.72 ± 1.09	1.42 ± 0.60
ASMOF	3.00 ± 2.78	2.69 ± 2.44
ASM*	2.75 ± 2.58	1.91 ± 0.72

Table 5.8: (a) Overlap coefficient Ω (%) and (b) distance measure Δ (pixels) between the manual and automated chest radiograph lung field segmentation. Various automated schemes were evaluated as indicated. The mean value over all images in the validation set and its standard deviation are reported.

Algorithm	Overlap (%)		Intensity energy $E_f(\mathbf{l})$	
	$n = 14$	$n = 40$	$n = 14$	$n = 40$
GT	100.0	100.0	1.19	1.11
MISCG	91.2	93.0	0.90	0.88
GTPCA	90.7	93.5	1.60	1.21
ASM*GT	89.8	91.2	1.23	1.07
ASM*	86.8	89.8	1.33	1.17

Table 5.9: Explains why MISCG performs better than ASM for lung field (left + right) segmentation in chest radiographs. Overlap values are reported for GT (ground-truth), MISCG, GTPCA (fit of GT in PCA subspace), ASM*GT (ASM* initialized with GT) and ASM*. The second column contains the ASM* intensity cost $E_f(\mathbf{l})$ for the resulting contours.

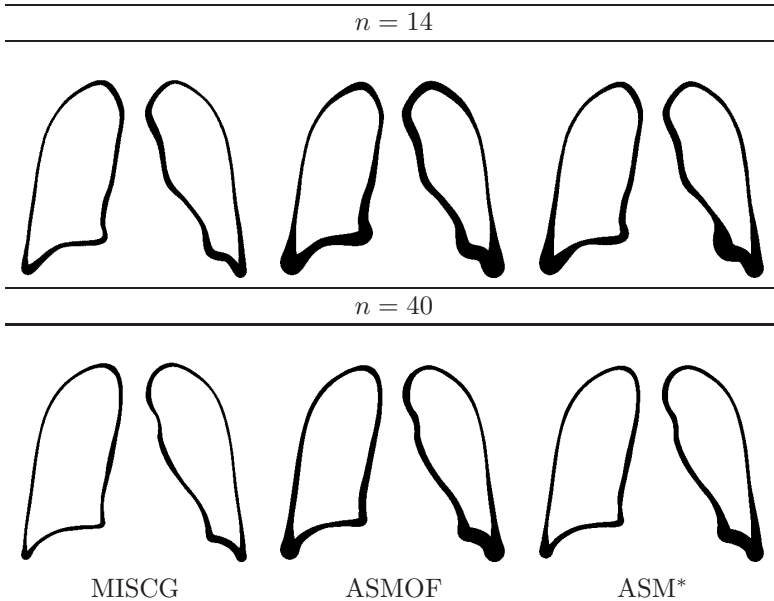


Figure 5.16: Local variation of the distance error along the lung field contours generated by the MISCG, ASMOF and ASM* algorithm. The mean error along the curve ranges from 1.3 to 9.1 pixels for ASM* and from 1.0 to 4.9 pixels for MISCG if 14 landmarks are used. With 40 landmarks, the errors decrease to 1.0 – 7.6 pixels for ASM* and to 0.8 – 4.3 pixels for MISCG. The ASM methods are less accurate at the lower regions (at the mediastinum) than MISCG.

5.7 Conclusion

The segmentation of an object from an image requires a model of the object and an algorithm to fit the model to the image. Both the construction of the model and the model fitting require a number of choices to be made. This can be a parameter to be set, but also a methodological decision to be made. In this chapter, insight is gained in all these choices by performing an in-depth validation on lung field segmentation from chest radiographs. Each part of the method was analyzed carefully to find appropriate parameter values and to explain the methodological choices to be made.

First, a proper object representation is required. One must decide where on the shape landmark points should be placed and where the connections should be defined. As many anatomical objects have visible borders and do not have very useful appearance information at the inside, landmarks are mostly placed at the object boundaries. The edges should be defined between landmark points that directly interact. For objects that have similar orientations and scales, it is reasonable to define the edges between direct neighbors along the object boundary.

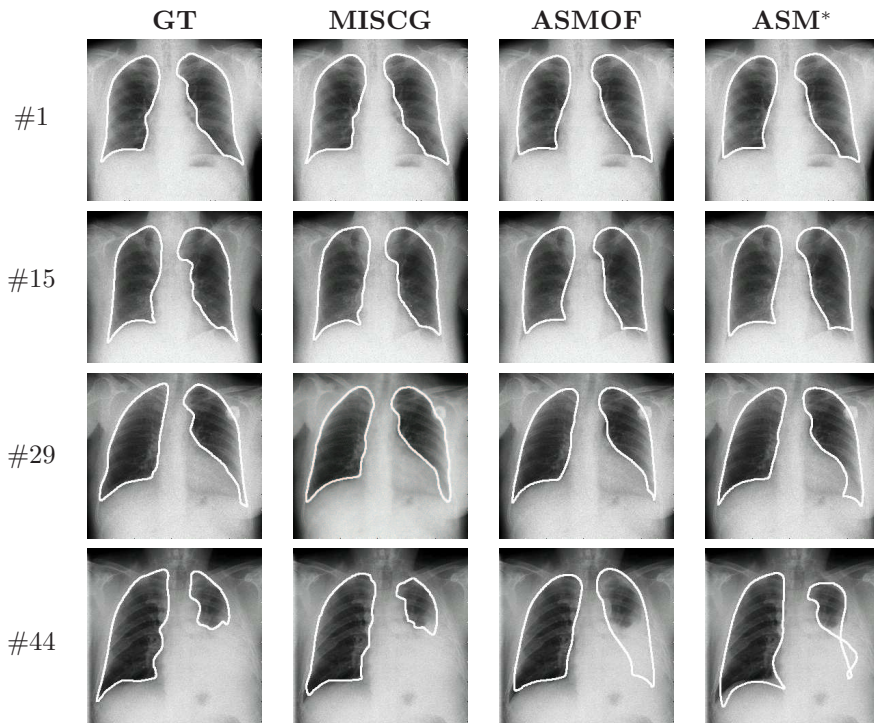


Figure 5.17: Segmentation results of the MISCG, the ASMOF and the ASM* algorithm on five validation images ranked from easy to difficult. The difficulty ranking was achieved as follows. For each image the overlap measure of the three algorithms with respect to the manual segmentations was averaged. The images were ranked according to the mean overlap from high to low. The images at the positions 1, 15, 29 and 44 were selected. The ground-truth (GT) segmentations are added for comparison.

Secondly, the intensity model requires several parameters to be specified. A first parameter series determines how the LOIs are computed from the original image. If a large set of LOIs is computed, probably only a limited set of them will significantly contribute in the detection of landmark candidates. However, as no (or only limited) segmentation performance can be gained by reducing the LOIs to only a limited set of optimal features and also because an exhaustive procedure is needed to find them, we prefer to compute an extensive set of LOIs and to use them all during segmentation.

The spherical profile configuration is specified by a length (number of sampled points) and a radius. Experiments have shown that the choice for the number of samples is not critical. The choice of the profile radius however, has a significant influence on segmentation performance. The radius controls the extent of the local image neighborhood around the landmarks. A descriptor with a

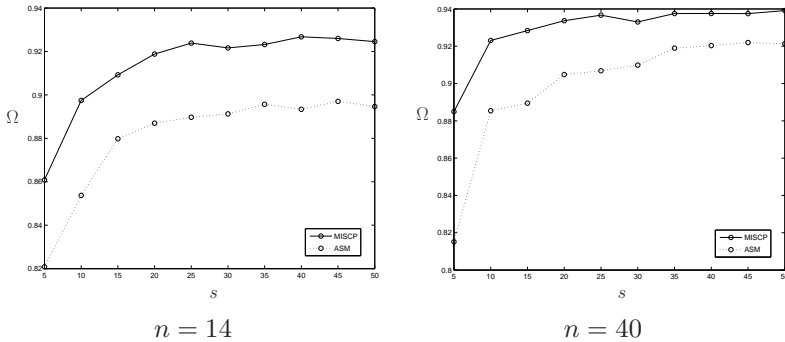


Figure 5.18: Influence of the number of training images on the segmentation performance. The overlap scores for the right lung with 14 and 40 landmarks are reported.

larger radius captures information about a larger part of the image whereas a smaller profile looks more locally at the image data. Consequently, a sufficiently small profile enables to localize the landmark positions more accurately but is more likely to generate false positives than a larger profile. However, despite their lower robustness, the smaller profiles perform significantly better for segmentation than the larger ones. The use of landmark-individual search regions and the incorporated shape constraints into the energy function avoid the algorithm of selecting wrongly positioned candidates.

A last parameter related to the intensity model was previously introduced to bring the intensity energy into a statistical meaningful range. Apparently, according to a number of experiments, this parameter can be estimated correctly by bringing the intensity and shape energies of the segmented shapes into balance.

The optimal segmentation scheme is one with a sufficiently fine search grid, an accurate profile and a sufficiently large number of selected candidates. However, such a scheme is not very efficient from a computational point of view. Instead, the same accuracy but with less computation time can be obtained using a multi-resolution strategy. The algorithm is started with a large spacing of the search points and sufficiently large search regions. In the following steps, smaller regions are sampled with a smaller spacing and more accurate profiles. This way, the number of evaluations of the intensity energy and the required number of candidates is reduced without losing quality.

Finally, a discrete optimization algorithm must be chosen to solve the combinatorial problem. Despite the fact that DP always finds the global optimum, it can not be used in every situation. With DP, computation time becomes easily unacceptable when dealing with complex graph configurations. If the computation of a path ($|\mathcal{A}| = 1$) would take one single second, the computation of a graph with $|\mathcal{A}| = 3$ would need almost three hours if $m = 100$ candidates are selected. In these situations MFA seems a good alternative as

its complexity does not suffer from the graph configuration. However, MFA often needs a lot of iterations and therefore the alternative IDP search method should be considered as a faster alternative. For lung field segmentation, no significant difference in segmentation accuracy between the three algorithms could be observed.

In this thesis, the choice for using local instead of global models was motivated. It was claimed that these models would allow to fit more accurately to unseen objects. This was empirically verified by comparing the proposed segmentation scheme (MISCG) against ASM for lung field segmentation. The better results obtained with MISCG could be attributed to a combination of two reasons. First, ASM segmentation is prone to get stuck in a wrong local optimum whereas MISCG, in the case of contour representations, finds a global optimum. Secondly, although the local shape models allow unrealistic shapes they indeed have the ability to fit more accurately to unseen images.

Chapter 6

Applications

6.1 Introduction

Object segmentation is essential in medical image processing. The extraction of the shape of anatomical objects from medical images has many purposes. It is often a preprocessing step towards another task. For example, a computer program that extracts lung nodules from chest radiographs, will first delineate the lung fields prior to searching for nodule candidates. Segmentation is also required for the automation of anatomical measurements and it is a possible means towards diagnosis, therapy planning and visualization. Hence, it is obvious that many applications exist for the proposed method.

The medical image segmentation task is both a two- as a three-dimensional problem. In radiography, an image is acquired by projecting X-rays on a 2D plate. Consequently, the radiographic image only shows two-dimensional projections of true three-dimensional shapes. The delineation of these shapes is often a challenge due to a large shape variability that originates from both population as pose differences. Also the overlapping projections of other anatomical structures makes the task more complex. The MISCG method will be demonstrated for the following radiographic applications: lung fields, heart and clavicles from chest radiographs (Section 6.2), delineation of the hand bones (Section 6.4) and segmentation of the femur and tibia in knee radiographs (Section 6.5). Also a clinical experiment was set up to validate the automated assessment of the cardiothoracic ratio from chest radiographs (Section 6.3). Finally, the MISCG method was applied for 3D liver segmentation from computed tomography scans (Section 6.6).

6.2 Anatomical structures in chest radiographs

The existence of many segmentation algorithms that can handle one single application requires comparative studies between these different methods. When a new method is published, it is compared against conventional methods. How-

ever, these comparisons are often not very trustworthy because of several reasons. Often the presented algorithms are developed for one particular database and are therefore optimally tuned for that particular database. Also more attention is given to the implementation of the presented method, than to the implementation of the conventional method. This difficulty has motivated the research community to come up with comparative studies using publicly available databases. The SCR database [11, 12] is one of these initiatives and was set up for the segmentation of anatomical structures in chest radiographs. We used this database to compare our MISC method to other approaches.

6.2.1 Data

The database consists of 247 standard posterior-anterior chest radiographs taken from the JSRT database [13]. The radiographs originate from 13 institutions in Japan and one in the United states. The images contained 2048×2048 pixels and a spatial resolution of 0.175mm per pixel and 12 bit gray-levels. In each image, the left and right lung field (Fig. 6.1(a)), the heart and the left and right clavicle were delineated manually by two experts. It is important that these experts use a consistent definition of what is meant by the delineation of each anatomical structure. For example the lung fields are defined as those pixels for which radiation passed through the lungs but not through the heart, the mediastinum and the aorta.

The same approach as described in Chapter 5 was used to annotate the contours with landmark points: 50 and 44 landmarks for respectively left and right lung, the heart was described with 27 and the clavicles with 23 landmark points.

The segmentations of the first expert are considered as the gold standard. The computer outcomes and the segmentations of the second observer are compared to the gold standard by means of the overlap Ω (Eq. 5.3) and closest distance error Δ (Section 5.6). The ultimate goal for a computer algorithm is to obtain an error that is comparable or even lower than the error¹ made by the second observer.

The images were divided in two parts by selecting the even and odd numbers. First, the images in one set are used to train the algorithm that is subsequently applied to the images of the other set. Second, a model is built from the second set and applied to the first set.

6.2.2 Methods

The segmentation methods that participated in this study all belonged to one of the following categories: active shape models, active appearance models and pixel classification. For a complete overview of all methods we refer to [11]. Only a short summary of the here reported methods is given.

¹Error is not the correct word as the segmentation of the second observer is not necessarily worse than the gold standard. The error mentioned here expresses the inter-observer variability.

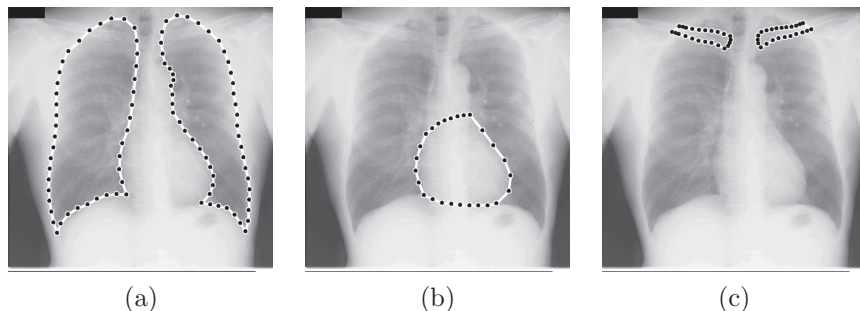


Figure 6.1: Anatomical structures to be segmented from one chest radiograph: lung fields (a), heart (b) and clavicles (c).

ASM tuned

The standard ASM scheme as described in Section 2.4.1 with the appearance model put forward in [54] was applied with optimal settings obtained from a number of pilot experiments.

AAM whiskers BFGS

Some alterations are made to the basic AAM algorithm (Section 2.4.2). As the standard AAM scheme only considers the object's interior for constructing the texture vectors, and because the interior of the anatomical objects often does not discriminate information, the search is likely to result in a segmentation that lies completely inside the object. Therefore, the model was augmented with texture information about the contour edges by adding intensities sampled along the contour normals at each landmark point.

A second alteration to the original AAM was that an extra refinement of the model fit was performed using quasi-Newton optimization algorithm. The abbreviation BFGS (Broyden, Fletcher, Goldfarb and Shanno) refers to the used update of the Hessian.

Pixel classification post-processed

The used pixel classification algorithm classifies each pixel according to three kinds of features: pixel position (x and y), the gray-level value of that pixel and the output of several gaussian derivative filters. A kNN classifier was trained with $k = 15$. During segmentation, the classifier was used to compute the probability that a pixel belongs to a specific class. To ensure that for every structure a single connected object was obtained, a post-processing procedure was applied. First, the soft output was blurred, then the largest connected object was selected and the holes in the object were filled.

MISCG

For MISCG, a model was constructed for each individual structure individually. The same LOIs were computed as for the experiments in Chapter 5. Profiles of length $n_f = 4$ and radii ranging from 8 to 4 pixels were extracted. Model fitting was performed with $m = 100$ candidates and DPCC was used for optimization.

6.2.3 Results

The overlap scores of each segmentation scheme on each anatomical object are reported in Table 6.1. More details and individual segmentations can be found on the website of the study [12]. The MISCG method seems to have the highest performance for each anatomical object. The lung fields are segmented with an accuracy comparable to manual segmentation. For the delineation of the heart and the clavicles, the second observer still scored significantly better. The outlines of the heart and the clavicles are often difficult to perceive in a radiographic image. Nevertheless, the second observer still segments these structures correctly as he is able to use contextual information. The recognition of other overlapping or neighboring structures helps him during segmentation. The automated algorithms on the other hand only use information about one single structure.

The overall result is shown in Fig. 6.2 where the pixel error is shown. This error is defined as the proportion of pixels for which any of the five object labels (right lung, left lung, heart, right clavicle, left clavicle) is not in agreement with the reference standard. MISCG has the lowest mean pixel error (0.033) except for the human observer (0.029), compared to 0.043 for pixel classification. Methods that use a global PDM as ASM tuned and AAM whiskers BFGS score worse with respectively 0.044 en 0.046. The main reason for this is the lack of flexibility of the global PDMs. We refer to the SCR website [12] for a detailed overview of the performance of all available methods described in [11].

6.3 Cardiothoracic ratio

One of the clinical motivations for lung field segmentation is that it enables the computation of the cardiothoracic ratio (CTR), that expresses the size of the heart relative to the size of the thorax as shown in Fig. 6.3. It is computed as

$$\text{CTR} = \frac{\text{MRD} + \text{MLD}}{\text{ID}}$$

where MRD and MLD denote the maximum diameter from the centerline to respectively the right and left heart border and ID stands for internal diameter (of the chest). A ratio of more than 0.5 is considered abnormal in an adult. An enlarged heart might be an indication of cardiac failure or pericardial effusion. The CTR is extracted from the segmented contours using a number of specific

	$\mu \pm \sigma$	Min	Q1	Median	Q3	Max
<i>Lungs</i>						
MISCG	95.3 ± 02.6	83.4	94.7	95.8	96.4	97.7
PC post-processed	94.5 ± 02.2	82.3	93.9	95.1	95.8	97.2
Human observer	94.6 ± 01.8	82.2	93.9	94.9	95.8	97.2
ASM tuned	92.7 ± 03.2	74.5	91.7	93.6	94.6	94.6
AAM whiskers BFGS	92.2 ± 02.9	71.8	91.4	93.1	94.0	96.1
<i>Heart</i>						
Human observer	87.8 ± 05.4	57.1	84.3	88.8	91.6	96.5
MISCG	84.9 ± 10.4	26.1	82.4	87.4	90.7	96.3
AAM whiskers BFGS	83.4 ± 07.0	51.0	79.1	84.5	88.2	96.7
PC post-processed	82.4 ± 07.7	50.0	78.3	84.4	87.7	93.2
ASM tuned	81.4 ± 07.6	52.0	77.0	82.7	87.3	93.8
<i>Clavicles</i>						
Human observer	89.6 ± 03.7	70.7	88.0	90.5	92.2	95.2
MISCG	77.5 ± 12.9	01.1	73.8	80.7	85.4	91.5
ASM tuned	73.4 ± 13.7	09.3	70.5	77.6	82.2	91.2
AAM whiskers BFGS	64.2 ± 17.1	00.3	58.8	68.9	76.1	86.1
PC post-processed	61.5 ± 12.3	22.3	55.4	63.9	70.6	83.7

Table 6.1: Overlap (%) scores for each anatomical structure. The minimum, maximum, median and first and third quartile are listed. The systems are ranked according to the median as in [11].

landmark points as shown in Fig. 6.3. The center line is positioned at the middle of landmarks **a** and **b**. The heart diameter is determined as the sum of at one hand the distance from the most left point between **c** and **d** to the centerline and at the other hand the distance from the most right point between **e** and **f** to the centerline. Similarly, the diameter of the thorax is computed from **g**, **h**, **i** and **j**.

The CTR was computed three times for every image. First, it was obtained by a radiologist manually as he does in his daily practice (R_1). Secondly, the CTR was computed from the manually delineated lung fields (R_0). A last result was obtained fully automatically (R_2). One case was left out as no complete left lung was available for one patient. The ratios computed from the manually delineated contours were considered as ground-truth. All values ranged from 0.37 to 0.70. It was found that 19 out of 43 patients had a CTR higher than 0.5 which could be attributed to the mix of normal and abnormal cases. The errors made by the automated and manual technique are summarized in Table 6.2. A paired t -test gave no significant difference between the manually and automatically obtained ratio. A Bland and Altman plot (Fig. 6.4) confirmed that the automated ratio gave errors comparable to the manual errors.

In a next experiment, the automatic computation of the lung fields and the

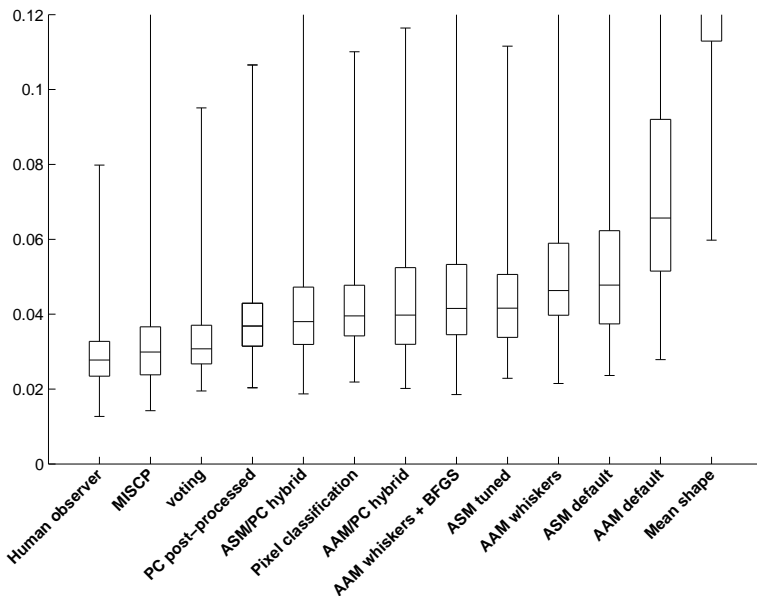


Figure 6.2: SCR database results: box plot (minimum, maximum, median and first and third quartile) of the pixel error of all available algorithms, sorted by increasing median error. The proposed method (MISCP where P stands for path) scores best of all automated approaches in this study.

CTR was implemented in the clinical workflow of the Radiology Department of the University Hospital Gasthuisberg in Leuven (Belgium) in collaboration with Prof. Dr. Verschakelen. Each time a radiographic image was acquired, the lung fields were delineated automatically and the result was stored temporarily. When a radiologist handled a case, he could press a CTR-button that enabled him to view the result. He was also allowed to give a score from 1 to 5 concerning the quality varying from completely wrong to very good. The average score was 4.0 and the median was found to be 5.

	$\mu \pm \sigma$	Min	Q1	Median	Q3	Max
Man	0.019 ± 0.020	0.000	0.005	0.014	0.024	0.089
Auto	0.025 ± 0.023	0.000	0.008	0.018	0.039	0.092

Table 6.2: Absolute errors made by the automated and manual technique to determine the CTR from a chest radiograph. A student *t*-test gave no significant difference between the manually and automatically obtained ratio.

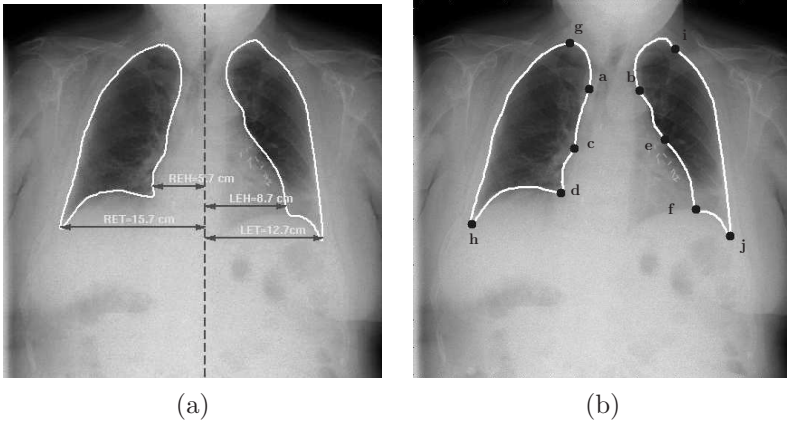


Figure 6.3: The cardiothoracic index or ratio measures the size of the heart relative to the size of the thorax. It is computed as the ratio between heart diameter (left part + right part) to the diameter of the thorax.

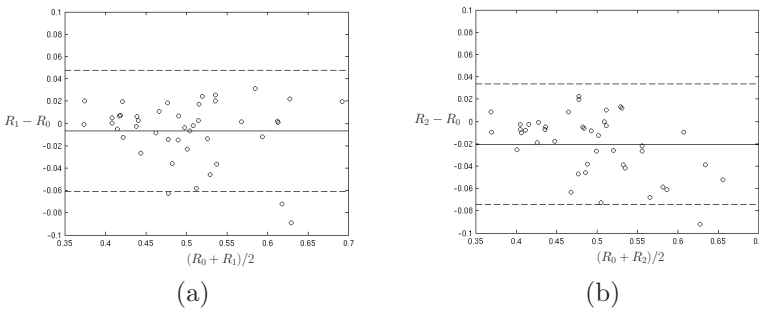


Figure 6.4: Bland and Altman plot for (a) the manual method and (b) the automated method. The difference between two measurements (X -axis) is plotted versus the average of the two measurements (Y -axis). According to this plot, the error of the automatic method (R_2) is comparable to the error of the manual method (R_1).

6.4 Hand bones

In order to demonstrate the generic nature of the algorithm, it was also applied for the segmentation of hand bones from a hand radiograph.

6.4.1 Data

The database consisted of 186 hand radiographs (image width of 256 pixels and varying height) and manual delineations of the bones of the little finger, the middle finger and the thumb. Each individual bone was described as a contour consisting of 20 (smaller ones) or 40 (bigger ones) landmarks (Fig. 3.2(f)). The

dataset is split in two folds, 93 images each. The first one is used for model building, the second one for validation. Due to the image acquisition protocol, all hands had a similar orientation which made an affine warping prior to model building unnecessary.

6.4.2 Model construction

One strategy is to represent each bone with an individual contour. However, as the locations of the bones are mutually correlated, it would be better to consider all the bones together as one single object by defining edges between the different bone contours.

Alternatively, a hierarchical model can be built. A first model is used to localize the center of each bone. This model is trained with landmarks at the bone centers and edges between the correlated centers. A second layer of models is needed to delineate each bone individually in a region centered at the corresponding bone center.

Two graphs were tested for the localization of the bone centers. The first one (Fig. 6.5(a)) models the hand as if the locations of the three finger tops were independent. A second model (Fig. 6.5(b)) also learns the relationship between finger-to-finger connections as shown in Fig. 6.5(b). The profiles were configured with $n_f = 4$ samples and two radii: $r_f = 8$ and $r_f = 4$ pixels. Again the same LOI sequence as in Chapter 5 was computed. An appropriate ρ was determined by the approach explained in Chapter 5 that brings the shape and intensity energy into balance.

For each of the eleven bones, a contour model is trained from sub-images extracted from the original training images as shown in Fig. 6.5(c). These sub-images are centered at the bone center and the range is chosen large enough to ensure that each of the 93 training shapes were fully captured. Intensity profiles for the contour model were extracted with $n_f = 4$ and $r_f = 1$ pixel.

6.4.3 Model fitting

The hierarchical model was fitted to the images in two consecutive steps. First, the bone-center model was applied to find proper locations for each bone in the image. The search occurred in two resolution levels: a coarse search with $r_f = 8$ and $\delta = 3$ pixels and a fine search with $r_f = 4$ and $\delta = 1$. A number of pilot experiments showed that selecting $m = 100$ candidates per landmark was sufficient. To find out which of the two proposed graphs should be applied on the validation set, a leave-one-out experiment was run on the training images with both graph configurations. Optimization was done with DP for the first graph (Fig. 6.5(a)) and MFA for the second graph (Fig. 6.5(b)).

To find the 11 bone contours, sub-images were extracted at the detected bone centers (Fig. 6.5(c)). The contour models were fitted using the DPCC algorithm with $m = 100$ and $\delta = 1$ pixel.

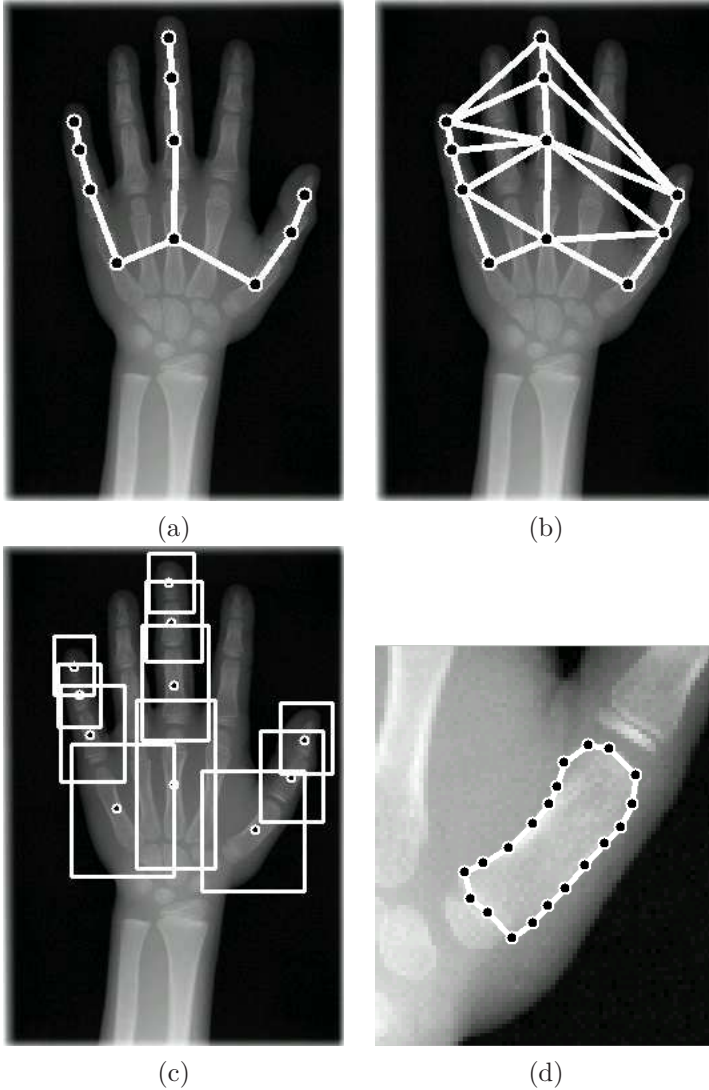


Figure 6.5: Segmentation strategy using a hierarchical object representation: the models (a) and (b) are built to localize the bone centers. Bone-individual contour models (d) are then applied to ROIs centered around the obtained centers (c).

6.4.4 Results

The leave-one-out experiment on the training images with the two bone-localization models gave the following results. The second model (Fig 6.5(b)) segmented 7 hands out of 93 wrongly. One error was made on a hand that had maximally opened fingers whereas the other training examples all had closed

or moderately opened fingers. The six other errors were due to a bad acquisition: bones of the little finger or the thumb lied too close to the image border. Such a bone is hard to detect as the corresponding intensity structure deviates too much from the model. These errors could be avoided by extending the image with zero pixels at the borders. However no border handling was used for all the results reported here.

The first model did worse with 11 wrong segmentations: the same 7 cases where the second model went wrong together with 4 additional erroneous cases. One of them is shown in Fig. 6.6. Obviously for this example, the incorporation of knowledge about the finger-to-finger correlation is needed to identify each bone correctly.

Hence, the 93 validation images were segmented by first applying the graphical model of Fig. 6.5(b) and subsequently delineating the bone contours. The bone localization went wrong in four cases due to bad acquisition. This way, an average overlap score of $\Omega = 90.15\%$ and distance error $\Lambda = 0.9$ pixels were obtained. These values improved to $\Omega = 91.5\%$ and $\Lambda = 0.5$ pixels without these four cases. An automated segmentation result is shown in Fig. 6.7(a). This case was selected as the one in the middle after ranking the images according to their overlap with the ground-truth Fig. 6.7(b). The overlap of this example was $\Omega = 92.4\%$.

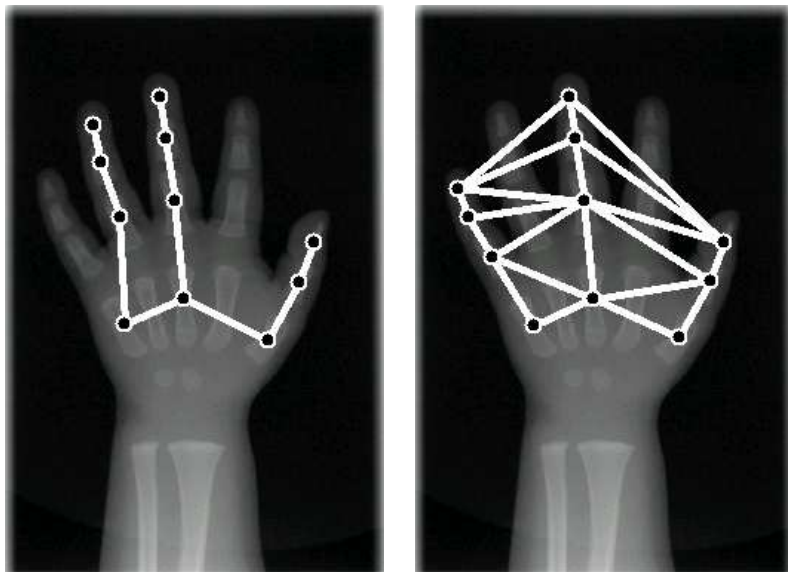


Figure 6.6: Example segmentation where the incorporation of knowledge about the finger-to-finger correlation is needed to identify each bone correctly.

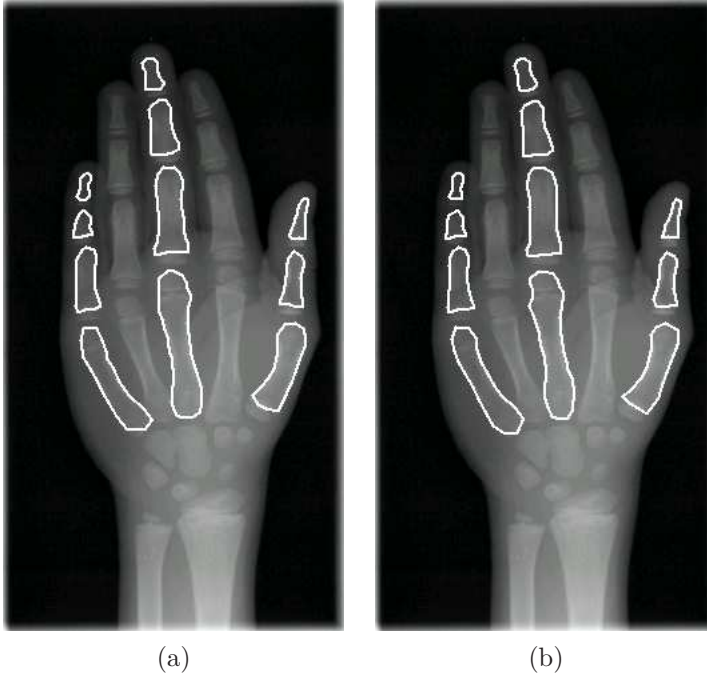


Figure 6.7: An automated segmentation result is shown in (a). This case was selected as the one in the middle after ranking the images according to their overlap with the ground-truth shapes (b). The overlap of this example was $\Omega = 92.4\%$.

6.5 Knee segmentation

In the clinical practice, radiography is frequently used for the diagnosis of knee osteoarthritis (OA). Arthritis refers to protective tissue (cartilage) that wears away. As a result, the distance between femur (upper bone) and tibia (lower bone) decreases. Consequently, a tool to automatically delineate the femur and tibia boundaries is a requirement towards the automated assessment of OA.

6.5.1 Data

A model was learnt from 38 example knee radiographs with expert manual delineations. One single object was used to represent femur and tibia together: a closed contour consisting of 39 landmark points (in Fig 6.8). Correspondences were obtained by first marking 6 specific anatomical landmarks and subsequently sampling pseudo-landmarks equidistantly in between them. The original images had a width of 1024 pixels and were resized for processing to be 512 pixels width (256 pixels for each leg). The analysis was performed only on the left knee and thus the images had a size of 256 pixels.

6.5.2 Experiments

A contour model was constructed with the following parameter values: $n_{\mathbf{f}} = 4$, $r_{\mathbf{f}} = 4$ pixels. The model was fitted to the training images (after leaving them out of the model) using the DPCC algorithm with $m = 100$ and three resolution levels: $\delta = 4, 2$ and 1 pixel.

6.5.3 Results

The automated delineations of femur and tibia were compared against the manual delineations using the closest distance error Λ . To ignore the influence of the error along the meaningless vertical connections between femur and tibia, the distance error was only measured along the bone boundaries. The results are summarized in Table 6.3 and three automated segmentations are shown in Fig. 6.8: (a) best score, (b) median score and (c) worst. The large error for (c) is 6.48 pixels and is caused by a pathology.

$\mu \pm \sigma$	Min	Q1	Median	Q3	Max
1.85 ± 0.96	0.80	1.33	1.58	2.08	6.48

Table 6.3: Knee application: closest distance error (pixels) made by the automated segmentation as compared to the ground-truth. The reported errors are the average errors along the bone boundaries.

6.6 Liver segmentation

The fact that liver segmentation from CT images is such a popular topic in medical image analysis (e.g. [18, 64, 77]), can be explained by two facts. Firstly, it is extremely relevant for clinical purposes. Liver segmentation is the basis for computer-based surgery planning and it can serve as a preprocessing step for diagnosis and monitoring. Secondly, automated segmentation is a highly challenging task because of a number of reasons. The border between liver and its surrounding tissue is not always clearly visible as illustrated in Fig. 6.9. The intensity range at the heart region (Fig. 6.9(a)) has some overlap with the range of liver tissue. The vena cava, a vessel fully surrounded by liver tissue and considered as part of the liver (Fig. 6.9(b)) is also a source of problems. Consequently, voxel classification according to intensity values alone is not possible. Also the high variability of the liver shape makes liver segmentation a challenge.

The performance of the MISCg method for liver segmentation was tested on a publicly available database [78]. A competition was started as part of the workshop *3D Segmentation in the Clinic: A Grand Challenge* [87] organized during the MICCAI conference in 2007.

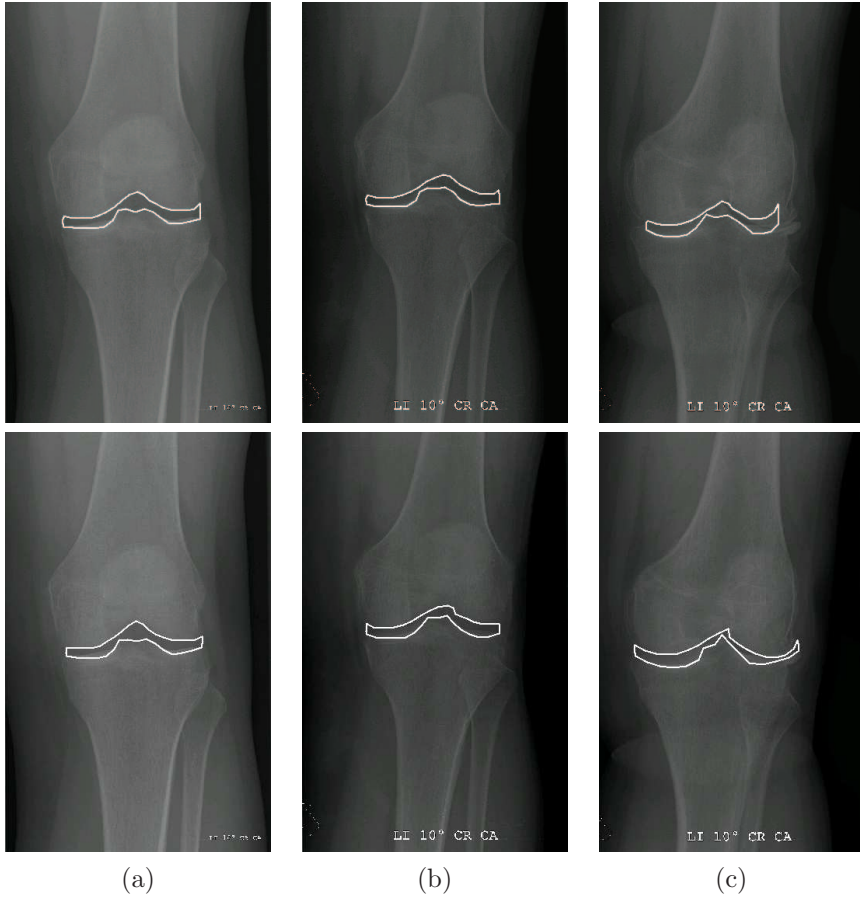


Figure 6.8: Three segmentation results: (a) smallest, (b) median and (c) largest error. The large error for (c) is 6.48 pixels and is caused by a pathologic knee. The automated segmentations are shown at the first row, the ground-truth at the second row.

6.6.1 Data

A set of 20 contrast enhanced CT datasets and corresponding segmentations of the liver were obtained from [78]. The images originated from several clinical institutes using different CT scanners for acquisition. Most images were pathologic as they included tumors, metastasis and cysts. All images were enhanced with contrast agent and scanned in the central venous phase. The in-slice pixel spacing varied from 0.55 to 0.8 mm and the inter-slice distance varied from 1 to 3 mm.

Ground-truth segmentations were created manually by radiologic experts. All internal structures as vessels and tumors surrounded by liver tissue, were considered as part of the liver.

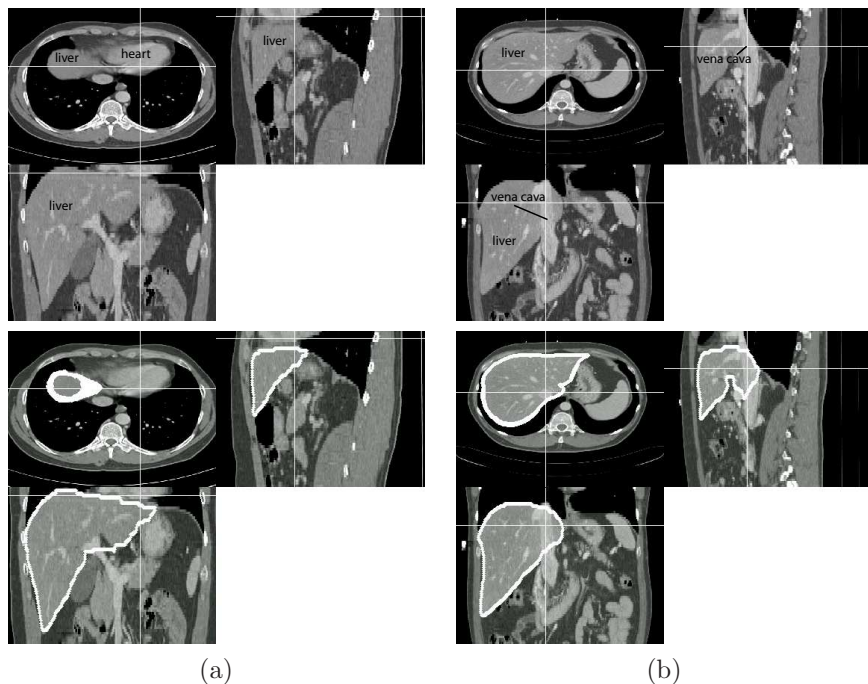


Figure 6.9: The border between liver and its surrounding tissue is not always clearly visible. The intensity range at the heart region (a) has some overlap with the range of liver tissue. The vena cava, a vessel fully surrounded by liver tissue and considered as part of the liver (b) is also a source of problems. The voxels that lie at the liver boundary are colored in white.

6.6.2 Model construction

The dataset as downloaded from [78] consisted of a set of images and a set of corresponding labeled images. A voxel had label 0 or 1 depending on whether the voxel lied outside or inside the liver. Consequently, prior to learning the shape and intensity models, several preprocessing steps were needed as to obtain a training set in the required format (images and landmark points).

Affine registration and resampling

One of the protocols used for scanning required patients to lie on their side. As to eliminate pose-related differences, an affine registration was needed. First, one image, referred to as the reference images, was arbitrarily chosen from the training set (first image of the training set). All other images and their corresponding segmentations were affinely transformed and sampled in the space of the reference image. More details about the registration can be found in [88]. The resulting resampled intensity images and segmentations all had a size of $256 \times 256 \times 185$ voxels with a voxel size of $1.5 \times 1.5 \times 1.5$ mm.

Morphological operations

Prior to extracting the liver surfaces, a morphological closing operation was performed to fill the holes and gaps in the label images. Holes and gaps up to 5 mm were closed this way.

Surface meshes

Next, the binary label images were converted to triangulated surface meshes using the Matlab FastRBF Toolbox [89]. The sides of the triangles had on average a length of 5 mm. The number of vertices ranged between 2573 and 4185. Some examples are shown in Fig. 6.10.

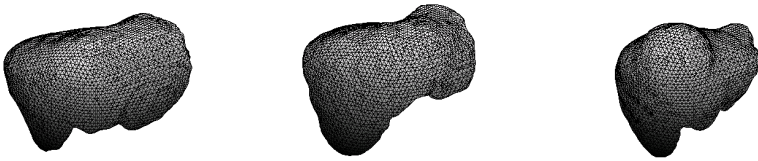


Figure 6.10: Three example livers represented as triangulated surface meshes.

Landmarks

Point correspondences between all training shapes were established using the point registration algorithm put forward by Claes *et al.* [80]. A short description of the algorithm was given earlier in Section 3.3.2. Some trial experiments where one reference liver shape \mathbf{r}_a was warped onto a target liver shape \mathbf{r}_b showed that still large errors could be observed between the warped shape $\mathbf{r}_c = T_{a \rightarrow b}(\mathbf{r}_a)$ and the target shape \mathbf{r}_b . A simple alteration made to the original algorithm (Fig. 6.11) made the error decrease significantly. Besides the warping from \mathbf{r}_a to \mathbf{r}_b (resulting in the shape \mathbf{r}_c) also a second warping from \mathbf{r}_b to \mathbf{r}_c was made. Finally, the inverse of the resulting transformation $T_{b \rightarrow c}$ was applied to \mathbf{r}_c . Hence, the final shape is computed as $\mathbf{r}_d = T_{b \rightarrow c}^{-1}(\mathbf{r}_c)$ with $\mathbf{r}_c = T_{a \rightarrow b}(\mathbf{r}_a)$. The shape \mathbf{r}_d has the configuration of \mathbf{r}_a but its points (approximately) lie at the surface of \mathbf{r}_b .

The algorithm as outlined above, was used to find corresponding landmarks on all shape instances. This was accomplished in two stages. In the first stage, one liver shape was chosen as a reference shape (first liver) and was deformed towards all remaining livers one by one. At this point, 20 shapes were obtained with an equal number of (corresponding) vertices. As to remove the bias in the landmark arrangement towards the chosen template shape, a second stage was carried out. Firstly, a new reference shape was obtained by averaging the 20 output shapes of the first stage. Secondly, this mean shape was used to compute a new mesh with equally sized triangles using FastRBF [89]. A

final set of landmark shapes with $n = 3277$ vertices and $t = 9825$ edges was obtained by registering the mean-shaped template towards all individual liver shapes.



Figure 6.11: A simple alteration made to the original algorithm [80]. Besides the warping from \mathbf{r}_a to \mathbf{r}_b (resulting in the shape \mathbf{r}_c) also a second warping from \mathbf{r}_b to \mathbf{r}_c was made. Finally, the inverse of the resulting transformation $T_{b \rightarrow c}$ was applied to \mathbf{r}_c .

Intensity model

A number of pilot experiments yielded a set of LOI parameters (Table 6.4) leading to 24 LOIs. For computational reasons the second order derivatives were not computed.

The spherical profiles were configured similarly as for the lung field experiments in Chapter 5: 2 samples are taken at each side of the center point in each direction, leading to a profile vector of length $n_f = 6$. Profile models with a larger ($r_f = 8$ pixels) and a smaller extent ($r_f = 4$ pixels) were built.

Symbol	Description	Value
Model construction		
n	number of landmarks	3277
t	number of edges	9825
n_f	profile length	6
r_f	profile radius	8, 4 pixels
σ	LOI inner scale	0.5, 1, 2 pixels
α	LOI outer scale	2σ pixels
L_i	LOI derivatives	L, L_x, L_y, L_z
M	LOI moments	1, 2
ρ	overestimation intensity energy	12
Model fitting		
δ_1	grid spacing (level 1, $r_f = 8$ pixels)	4 pixels
δ_2	grid spacing (level 2, $r_f = 4$ pixels)	2 pixels
δ_3	grid spacing (level 3, $r_f = 4$ pixels)	1 pixel
a	initial search grid extent	3
m	number of candidates	200

Table 6.4: Parameters for liver segmentation.

6.6.3 Model fitting

The model was fitted in three resolution levels (Table 6.4). At the first level, sufficiently large search regions were explored with a sampling interval $\delta_1 = 4$ pixels and the large-sized profile. The next resolution levels used the more accurate profile ($r_f = 4$ pixels) in smaller search regions centered around the landmark points that came out of the previous stage.

For optimization IDP was chosen because of its low computation time and a reduction rate of $f_{\text{red}} = 0.05$ was used. The number of selected candidates was set to $m = 200$, which was still acceptable from a computational point of view. In some cases it occurred that for a particular landmark point, mostly because of a pathologic region, no candidate was selected in the neighborhood of the true landmark location. These outlier landmarks could be detected after every optimization stage. If a landmark location \mathbf{l}_i resulted in an excessively high shape energy

$$\sum_{j \in \mathcal{N}_i} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) > 1000$$

this landmark was considered as an outlier and a new location was computed by averaging the neighboring landmark points. The needed computation time using non-optimized Matlab code with some computational subroutines written in C was approximately 1h on a 2.8 GHz Intel Pentium processor.

In a second experiment, the first two stages were carried out with the same settings as the previous experiment (using IDP) but the third resolution level was optimized with MFA.

6.6.4 Evaluation

The segmentation scheme was validated by segmenting the training images (using leave-one-out) and comparing the results against the reference standard. To enable a comparison between an automated segmentation (consisting of landmarks) and a ground-truth segmentation (label image), the landmark shapes were converted to binary label images: 0 if a voxel is background and 1 if the voxel lies inside the liver.

We applied the scoring system described in [87] which is based on five error metrics: volumetric overlap error, volume difference, surface distance, RMS surface distance and maximum surface distance. These errors were converted into scores between 0 and 100 by comparing the errors that the algorithm made to the errors that a second observer made when manually segmenting the same images. For a perfect result (error is zero), 100 points were awarded. For an error equal to the error of the independent human observer, a score of 75 was given. All other scores were computed by using a linear scale between these two gauge values. As to avoid negative scores, negative values were truncated to 0. Consequently, if an algorithm gets 75 points it is considered to perform as well as manual segmentation.

6.6.5 Results

The results are summarized in Table 6.5. The mean of every error is listed together with the corresponding score. Obviously, the algorithm still performs significantly worse than human with a mean score of 68.9%. However, it seemed that for 10 out of the 20 cases a score higher than 75% was obtained. The larger RMS distance error (2.9 mm) and maximum surface distance (28.0 mm) imply that the algorithm makes larger errors in some regions as compared to the second observer whereas in other regions a comparable accuracy is obtained. This is illustrated in Fig. 6.12 that shows the distance errors of the worst (#20), middle (#11) and best case (#1) with scores of 34.0%, 73.2% and 83.4% respectively. The results increased slightly (mean score of 69.5%) with the second segmentation scheme (final stage with MFA).

Considering on one hand the large shape and appearance variance of the liver, and the small number of examples on the other hand, it is very likely that a significant improvement could be made when the training set would be increased.

Measurement	Second observer		MISCG	
	error	score (%)	error	score (%)
Overlap error	6.4%	75	7.6%	70.1
Volume diff.	4.7%	75	2.9%	84.6
Avg. dist.	1.0 mm	75	1.3 mm	66.8
RMS dist.	1.8 mm	75	2.9 mm	59.7
Max. dist.	19 mm	75	28.0 mm	63.2
Total		75		68.9

Table 6.5: Liver segmentation results for the leave-one-out experiment. The mean of every error is listed together with the corresponding score. Obviously, the algorithm still performs significantly worse than the human observer (75%) with an overall score of 68.9%.

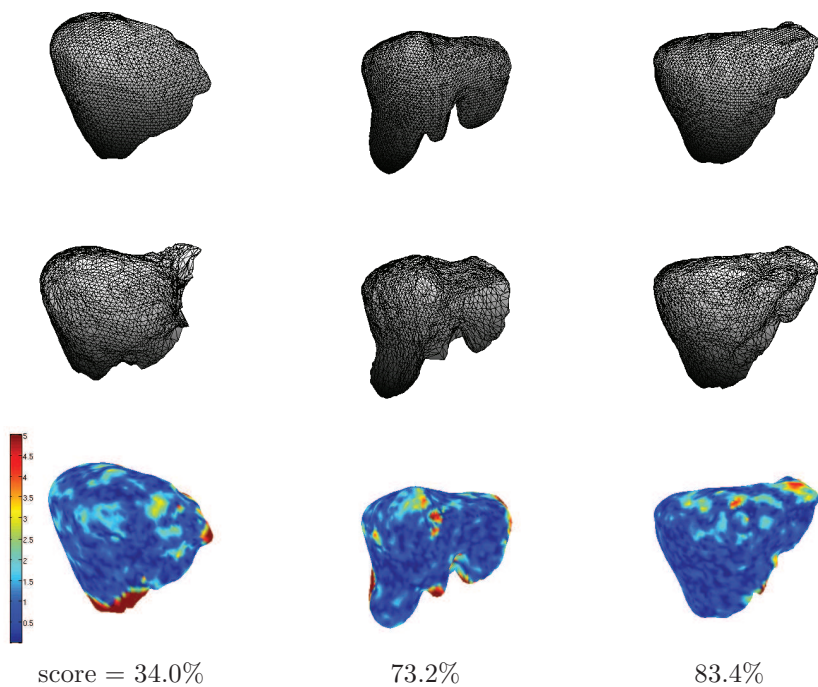


Figure 6.12: The worst (34.0%), middle (73.2%) and best (83.4%) segmentation obtained with MISC. Ground-truth and automated segmentations are shown in respectively the top and middle row. The colored shapes at the bottom row indicate the surface distance errors ranging from 0 mm (blue) to 5 mm (red).

Chapter 7

Conclusions and future work

7.1 Main contributions

The main goal of this thesis was to develop a model-based method for medical image segmentation that is able to fit accurately to unseen objects, even if only a limited number of training examples is available. Whereas most algorithms first build a global PDM by learning from a set of example images with corresponding landmark points and use this model to find instances of the modeled object in new images, a method was proposed that only considers local statistical dependencies. This approach has two major advantages. First, the model can generate accurate fits to new objects even if only a small number of training images is available. Second, an optimal or nearly optimal solution for the model fitting optimization problem can be found.

The method has a strong theoretical basis as it is formulated from a probability point of view. The intensities in the image and the shape of the object are considered as correlated random processes. This way of thinking enables to formulate the segmentation task as a maximum a posteriori problem: given an image, find the most likely shape. A key advantage of this approach is that a meaningful answer is given for the weighing between shape fitting and image appearance fitting.

The highly generic nature of the method was proven by applying it for the segmentation of several anatomical structures, both from 2D and 3D images. It was shown for multiple applications that the presented method outperforms other competitive methods.

7.2 Method

The object is represented as a graph that consists of a number of landmark points (vertices) defined on the shape and a number of connections between

neighboring landmarks (edges). The shape is modeled as a Markov random field such that the landmark points only have a (direct) statistical interaction with their neighbors. This allowed for the definition of a shape energy that reflects the plausibility of an unseen shape instance with respect to its distribution. This shape energy function is a sum of local energies that only depend on couples of neighboring landmark points.

The appearance model assumes that the object influences the local image structure around each landmark individually. The local intensity patterns are captured by extracting spherical intensity profiles centered at the landmark points from a number of feature images (LOIs) computed from the original image. By estimating statistical distributions for the LOI profiles, a global intensity energy could be defined that consists of landmark-individual energies that express the goodness-of-fit of the landmarks to the image.

Both models were merged into one object model and a corresponding object energy was defined that consists of first order (one vertex) intensity potentials and second order (two neighboring vertices) shape potentials. The intensity potentials drive the shape towards the correct image features whereas the shape potentials control the geometry of the shape. It was shown that finding the shape that minimizes this global object energy is equivalent to finding the most likely (MAP) shape in a given image.

As the objective function has many local optima and a huge amount of degrees of freedom, finding an optimum in the continuous domain with conventional techniques is like searching for a needle in a haystack. To cope with this, prior to optimization, the search space is discretized by selecting a number of candidate locations for each landmark individually.

The resulting combinatorial optimization problem behaves as an MRF and can be solved using several techniques. A first technique uses dynamic programming to find the global optimum. As the computation time of DP becomes easily unacceptable for graphs with complex configurations, two alternative techniques were proposed. Mean field annealing, an optimization method known in the world of MRFs, tries to minimize the energy of the MRF by estimating the mean field for decreasing Gibbs temperatures. A second alternative technique approaches the problem by iteratively solving minimal cost paths and removing the worst candidates.

As the computational complexity of the presented techniques increases with the number of candidates selected for each landmark, a search method was proposed that uses landmark-individual search regions and a multi-resolution strategy to keep the number of candidates as small as possible.

The method was initially developed and applied for lung field segmentation from chest radiographs. Insight, also transferrable to other anatomical objects, was gained in all the choices that had to be made during model construction and model fitting.

An important choice is where to define the vertices and edges on the shape. Anatomical objects typically have a clearly visible boundary which pleads for

landmark points on the object border. As the training shapes have similar orientations and scales, due to acquisition or due to an affine registration, it makes sense to define the edges only between the first neighbors along the object border.

Other choices concern the intensity model. A first parameter series determines how the LOIs are computed from the original image. If a large set of LOIs is computed, probably only a limited set of them will significantly contribute in the detection of landmark candidates. However, as no (or only limited) segmentation accuracy can be gained by reducing the LOIs to only a limited set of optimal features and also because an exhaustive procedure is needed to find them, we prefer to compute an extensive set of LOIs and to use them all during segmentation.

The spherical profile configuration is specified by a length (number of sampled points) and a radius. Experiments have shown that the choice for the number of samples is not critical. The choice of the profile radius however, has a significant influence on segmentation performance. A sufficiently small profile enables to localize the landmark positions more accurately than the larger ones but is more likely to generate false positive candidates. The use of landmark-individual search regions and the incorporated shape constraints into the energy function avoid the algorithm of selecting the false positives. Consequently, the smaller profiles, despite their lower robustness, perform significantly better for segmentation than the larger ones as long as enough candidates are selected.

A next set of parameter values specifies how the model is fitted to a new image. The optimal segmentation scheme is one with a sufficiently fine search grid, an accurate profile and a sufficiently large number of selected candidates. However, such a scheme is not very efficient from a computational point of view. Instead, the same accuracy but with less computation time can be obtained using a multi-resolution strategy. The algorithm is started with a large profile, a large spacing of the search points and sufficiently large search regions. In the following steps, smaller search regions are sampled with a smaller spacing and more accurate profiles. This way, the number of evaluations of the intensity energy and the required number of candidates is reduced without losing quality.

A final decision concerns the choice for an optimization algorithm. Despite the fact that dynamic programming theoretically finds the global optimum, it can not be used in any situation. With DP, computation time becomes easily unacceptable when dealing with complex graph configurations. In these situations mean field annealing seems a good alternative as its complexity does not suffer from the graph configuration. However, MFA often needs a lot of iterations and therefore the alternative IDP search method should be considered as a faster alternative.

7.3 Clinical relevance

If we look at the method from a clinical point of view, we must consider its possibilities and limitations in clinical practice. Can a computer algorithm really replace manual segmentation? For many applications, the computer has a human-like accuracy (e.g. CTR results in Section 6.3), but is more likely to fail as compared to manual segmentation when the anatomical structures show large pathological abnormalities. As these pathologies are often difficult to model, they complicate the computer's search for a correct segmentation. A human brain on the other hand is capable of recognizing these abnormalities and moreover, the brain will recognize and virtually delineate other neighboring anatomical structures. This pathological and contextual knowledge is very valuable for the manual segmentation process.

Hence, it is not advisable to use the automated segmentations blindly for further diagnostic purposes. Instead, an expert should still supervise the computer outcomes. Alternatively, instead of using a fully automated algorithm, an interactive computer program might be a good compromise. The user could for example specify the positions of a number of crucial landmark points.

7.4 Suggestions for future work

7.4.1 Curse of number of candidates

One of the prerequisites towards a correct segmentation is that for each landmark appropriate candidate locations are selected during the discretization step. When dealing with pathological cases, it occurs that a specific landmark does not really exist or is not visible in the image. Consequently, if only a limited number of candidates is selected, it might occur that the intensity detector will fail to select an appropriate location. For 2D applications, this seems not a problem, as the number of candidates can be chosen large enough such that each point in the search region is included in the candidate set. The shape energy in the objective function will force the algorithm to select a geometrically plausible candidate. This is illustrated in Fig. 7.1 for lung field segmentation. Selecting only 20 candidates per landmark was insufficient to find an appropriate point in the neighborhood of the implanted device. The problem was avoided by increasing the number of candidates.

However, when dealing with 3D objects with a large shape variability, large 3D search regions are scanned. In these cases, it is from a computational point of view unacceptable to drive up the number of candidates such that the whole search region is covered. We suggest a number of approaches to compensate for this.

Variable number of candidates

In its current implementation, the same number of candidates m is selected for each landmark individually. However, it might be a good idea to work with

a variable number m_i with $i = 1, \dots, n$. Select more candidates in delicate regions and a lower number for easily tractable landmarks. The challenge of this approach is to automatically determine (during segmentation) the required number of candidates for each landmark.

Outlier candidate

Another possible strategy is to add an extra *outlier* candidate for each landmark. The combinatorial problem of selecting one out of m candidates for each landmark becomes a problem of selecting one out of $m + 1$ candidates. The fictive outlier should be selected when no proper location is available for that landmark.

Combining local and global PDMs

Finally, an alternative strategy is to combine a global PDM with the proposed localized model. A shape \mathbf{I} could be considered as the sum of a global shape $\mathbf{I}_{\text{global}}$ that listens to a global PDM and a correction shape \mathbf{I}_{corr} that expresses the deviation of \mathbf{I} from the global PDM:

$$\mathbf{I} = \mathbf{I}_{\text{global}} + \mathbf{I}_{\text{corr}}$$

The correction shape \mathbf{I}_{corr} compensates for the lack of flexibility of the global PDM. The model is constructed in two stages:

1. First, a global PDM is constructed as explained in Section 2.4.1. This model will describe the global part $\mathbf{I}_{\text{global}}$.
2. Second, a localized PDM as described in this thesis is built to model \mathbf{I}_{corr} . This is obtained as follows. For each training shape $\mathbf{I}^{(i)}$ the following procedure is followed. First, a global PDM is built from all training shapes except $\mathbf{I}^{(i)}$. Second, the difference between $\mathbf{I}^{(i)}$ and its reconstruction $\mathbf{I}_{\text{rec}}^{(i)}$ using the reduced global PDM is computed. Repeating this for every training shape yields a set of local correction shapes:

$$\mathbf{I}_{\text{corr}}^{(i)}, \quad i = 1, \dots, s.$$

Finally, a localized PDM is built from the set of correction shapes.

During segmentation the global model is iterated using ASM (to find $\mathbf{I}_{\text{global}}$) and in each iteration the MISC method is applied to find the shape correction \mathbf{I}_{corr} . Consequently, only small search regions need to be explored and hence, only a limited number of candidates is needed.

7.4.2 Optimization algorithms

Three different optimization algorithms capable of finding optimal or near-optimal solutions of the discrete assignment problem were described. Very

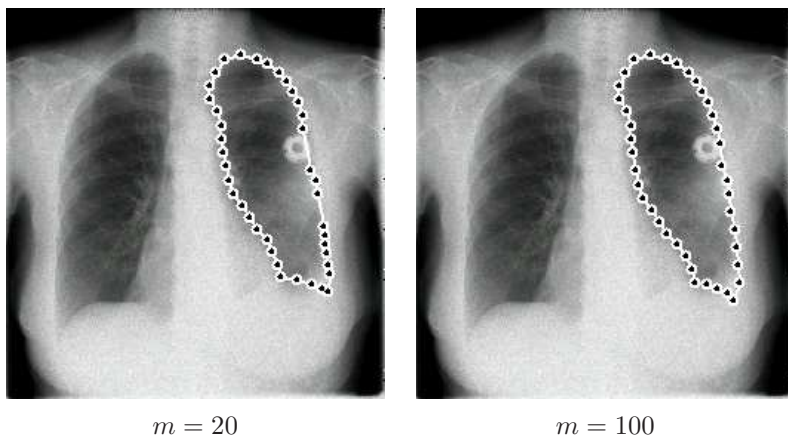


Figure 7.1: Selecting only 20 candidates per landmark was insufficient to find an appropriate point in the neighborhood of the implant. The problem was avoided by increasing the number of candidates to 100.

recently, there seems to have been an explosion of interest in such algorithms that solve the so-called *max-sum labeling problem* where the objective function is characterized by sums of functions of pairs of variables.

Recently published work [90, 91] seems strongly related to the proposed IDP algorithm. A comparison with these approaches would be very worthwhile. They have in common that the MRF energy is split up in easy-to-solve sub-problems (slaves). The slave-problems can be paths or trees as long as they are fast to compute. At the top level there is a master that iteratively coordinates the slave-individual solutions to make them gradually agree on common nodes in the overall MRF.

The main difference between IDP and for example the work of Komodakis [90] is the way the master does its coordination job. The IDP master does not change the slave MRFs but gradually prunes the set of possible labels per node. After a while, all slaves will select common labels. In Komodakis's work, the MRF energies of the trees are slightly changed at those places where the slaves do not agree such that gradually an agreement between the slaves is obtained. In the latter, the process of forcing-to-agree is less explicit than in the IDP method. Moreover, how the overall problem is split into sub-problems and how the sub-solutions are merged has a much stronger theoretical background.

7.4.3 Merging segmentation and registration

Registration and segmentation are often considered as intertwined and complementary problems in medical image analysis. Registration can be the solution for segmentation (atlas-based segmentation) and vice versa: if the same anatomical structures are segmented in both images, a correspondence between

the two images is established. Therefore, several researchers try to tackle both problems simultaneously (e.g. [92]). Also the framework described in this thesis is extendable for merging segmentation and registration.

Consider two images \mathbf{I}_A , \mathbf{I}_B of the same patient but acquired at different points in time. To find the segmentations \mathbf{l}_A and \mathbf{l}_B simultaneously, the following MAP problem needs to be solved:

$$(\mathbf{l}_A^*, \mathbf{l}_B^*) = \arg \max_{\mathbf{l}_A, \mathbf{l}_B} p(\mathbf{l}_A, \mathbf{l}_B | \mathbf{I}_A, \mathbf{I}_B) \quad (7.1)$$

$$= \arg \max_{\mathbf{l}_A, \mathbf{l}_B} \frac{p(\mathbf{l}_A, \mathbf{l}_B)p(\mathbf{I}_A, \mathbf{I}_B | \mathbf{l}_A, \mathbf{l}_B)}{p(\mathbf{I}_A, \mathbf{I}_B)} \quad (7.2)$$

The factor $p(\mathbf{l}_A, \mathbf{l}_B)$ in Eq. 7.2 expresses the likelihood of observing the shapes \mathbf{l}_A and \mathbf{l}_B in two images of the same patient:

$$p(\mathbf{l}_A, \mathbf{l}_B) = p(\mathbf{l}_A)p(\mathbf{l}_B | \mathbf{l}_A) \quad (7.3)$$

$$= p(\mathbf{l}_A)p(\mathbf{l}_B - \mathbf{l}_A | \mathbf{l}_A) \quad (7.4)$$

If we now make the simplification that the shape change does not depend on the value of the shape itself, then we have

$$p(\mathbf{l}_A, \mathbf{l}_B) = p(\mathbf{l}_A)p(\mathbf{l}_{AB}) \quad (7.5)$$

with $\mathbf{l}_{AB} = \mathbf{l}_B - \mathbf{l}_A$.

The factor $p(\mathbf{I}_A, \mathbf{I}_B | \mathbf{l}_A, \mathbf{l}_B)$ in Eq. 7.2 reflects the fitting of the shapes to the image data:

$$p(\mathbf{I}_A, \mathbf{I}_B | \mathbf{l}_A, \mathbf{l}_B) = p(\mathbf{I}_A | \mathbf{l}_A, \mathbf{l}_B)p(\mathbf{I}_B | \mathbf{l}_A, \mathbf{l}_B, \mathbf{I}_A) \quad (7.6)$$

$$= p(\mathbf{I}_A | \mathbf{l}_A)p(\mathbf{I}_B | \mathbf{l}_B, \mathbf{I}_A) \quad (7.7)$$

$$\approx p(\mathbf{I}_A | \mathbf{l}_A)p(\mathbf{I}_B | \mathbf{l}_B) \quad (7.8)$$

The conversion from Eq. 7.6 to Eq. 7.7 assumes that the intensities in one image only directly correlate with the shape in that image and not with the shape in the other image, which is highly acceptable. The next conversion is a rude simplification that assumes that the image structure around a shape does not depend on the image data of a previous image.

With Eq. 7.5 and Eq. 7.8, the MAP criterion of Eq. 7.2 can be rewritten as

$$(\mathbf{l}_A^*, \mathbf{l}_B^*) = \arg \max_{\mathbf{l}_A, \mathbf{l}_B} p(\mathbf{l}_A)p(\mathbf{l}_{AB})p(\mathbf{I}_A | \mathbf{l}_A)p(\mathbf{I}_B | \mathbf{l}_B) \quad (7.9)$$

with $p(\mathbf{I}_A | \mathbf{l}_A)$ and $p(\mathbf{I}_B | \mathbf{l}_B)$ computed with the intensity energy of Eq. 3.50. The shape energy of Eq. 3.38 enables the computation of $p(\mathbf{l}_A)$. The novel part in Eq. 7.9 is $p(\mathbf{l}_{AB})$ that reflects the plausibility of a change in shape over time. A model that describes the shape changes can be learnt from a training set that consists of couples of images of the same patient but acquired at different points in time.

Thanks to such an approach it might be expected that the simultaneous segmentations are more consistent in shape than if they were segmented individually. This technique would be useful for patient follow-up or in applications where several time frames of the same structure are acquired.

Appendix A

Conditional edge energy

If the second order approximation presented in Section 3.5.3 is made, only second order potentials (Eq. 3.41) are needed to compute the shape energy. If at the other hand the exact energy is used and if the graph consists of cliques larger than order two, also the higher order potentials need to be estimated. In that case, also the conditional edge distributions $p(\mathbf{e}_{ik}|\mathbf{e}_{ij})$ (Eq. 3.30) need to be estimated. Assume that the vector \mathbf{x} that holds the edges \mathbf{e}_{ij} and \mathbf{e}_{ik}

$$\mathbf{x} = \begin{bmatrix} \mathbf{e}_{ik} \\ \mathbf{e}_{ij} \end{bmatrix} \quad (\text{A.1})$$

is normally distributed:

$$\mathbf{x} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right) \quad (\text{A.2})$$

Also note that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_{\mathbf{e}_{ik}}$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_{\mathbf{e}_{ij}}$, $\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{\mathbf{e}_{ik}}$ and $\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{\mathbf{e}_{ij}}$. It can be shown [84] that the edge \mathbf{e}_{ik} for a given \mathbf{e}_{ij} is also normally distributed:

$$\mathbf{e}_{ik}|\mathbf{e}_{ij} \sim N(\boldsymbol{\mu}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}}, \boldsymbol{\Sigma}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}}) \quad (\text{A.3})$$

where the conditional mean depends on the value of \mathbf{e}_{ij} as follows (see [84]):

$$\boldsymbol{\mu}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}} = \boldsymbol{\mu}_{\mathbf{e}_{ik}} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{\mathbf{e}_{ij}}^{-1}(\mathbf{e}_{ij} - \boldsymbol{\mu}_{\mathbf{e}_{ij}}) \quad (\text{A.4})$$

The conditional covariance is a constant with respect to \mathbf{e}_{ik} :

$$\boldsymbol{\Sigma}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}} = \boldsymbol{\Sigma}_{\mathbf{e}_{ik}} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{\mathbf{e}_{ij}}^{-1}\boldsymbol{\Sigma}_{21} \quad (\text{A.5})$$

A formula for the conditional edge energy follows from combining Eq. 3.30 and Eq. A.3:

$$d_{ik|ij}(\mathbf{l}_i, \mathbf{l}_j, \mathbf{l}_k) = (\mathbf{l}_k - \mathbf{l}_i - \boldsymbol{\mu}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}})^T \boldsymbol{\Sigma}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}}^{-1} (\mathbf{l}_i - \mathbf{l}_k - \boldsymbol{\mu}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}}) \quad (\text{A.6})$$

with $\boldsymbol{\mu}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}}$ and $\boldsymbol{\Sigma}_{\mathbf{e}_{ik}|\mathbf{e}_{ij}}$ computed as in Eq. A.4 and Eq. A.5 respectively. In words: the conditional energy of an edge \mathbf{e}_{ik} for a given \mathbf{e}_{ij} is computed as the

Mahalanobis distance between this edge and its expectation value for a given \mathbf{e}_{ij} . The unknown distribution parameters in Eq. A.3 are estimated from the training shapes using the ML principle.

Appendix B

Expected value of the Mahalanobis distance

Assume a normally distributed random vector $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (\text{B.1})$$

with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The Mahalanobis distance d between an instance \mathbf{x} and its mean $\boldsymbol{\mu}$ is computed as

$$d(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (\text{B.2})$$

This distance measure expresses the goodness-of-fit of a sample according to its distribution. Two samples with the same Mahalanobis distance have the same likelihood according to the normal distribution, which means that they lie at the same ellipsoid centered around $\boldsymbol{\mu}$.

A disadvantage of the computation as in Eq. B.2 is that the inverse of the covariance matrix $\boldsymbol{\Sigma}$ is required. Consequently, if the covariance matrix is singular, Eq. B.2 becomes invalid.

A more general calculation of the Mahalanobis distance can be obtained from the eigenvalue decomposition of $\boldsymbol{\Sigma}$. As $\boldsymbol{\Sigma}$ is a real symmetric matrix, it has only real eigenvalues. If r denotes the rank of $\boldsymbol{\Sigma}$, then $\boldsymbol{\Sigma}$ has r nonzero eigenvalues $\lambda_1, \dots, \lambda_r$ with corresponding eigenvectors $\boldsymbol{\Phi}_1 \dots \boldsymbol{\Phi}_r$. This eigenvalue decomposition can be used to transform \mathbf{x} into a random vector \mathbf{z} with independent components with zero mean and unit variance:

$$\mathbf{z} = \begin{bmatrix} 1/\sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 1/\sqrt{\lambda_r} \end{bmatrix} \boldsymbol{\Phi}^T (\mathbf{x} - \boldsymbol{\mu}) \quad (\text{B.3})$$

with $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1 \dots \boldsymbol{\Phi}_r]$.

The Mahalanobis distance is now computed as

$$d(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = d(\mathbf{z}, 0, I_r) = \mathbf{z}^T \mathbf{z} \quad (\text{B.4})$$

and its expected value becomes

$$\langle d(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \rangle = \sum_{i=1}^r \langle z_i^2 \rangle = \sum_{i=1}^r 1 = r \quad (\text{B.5})$$

Bibliography

- [1] W. C. Röntgen. On a new kind of rays. *Nature*, 53:274–277, 1896. vii, 1
- [2] G. Hounsfield. Computerized transverse axial scanning (tomography). 1. description of system. *The British Journal of Radiology*, 46(552):1016–1022, 1973. vii, 1
- [3] P. Suetens. *Fundamentals of Medical Imaging*. Cambridge University Press, 2002. vii, 1, 6, 10
- [4] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–366, 1994. viii, 2, 12, 39
- [5] T. Heimann, S. Münzing, H.P. Meinzer, and I. Wolf. A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. In *Information Processing in Medical Imaging*, volume 4584 of *Lecture Notes in Computer Science*, pages 1–12, Kerkrade, The Netherlands, July 2007. ix, 2, 19, 20
- [6] S. Gleason, H. Sari-Sarraf, M.A. Abidi, O. Karakashian, and F. Morandi. A new deformable model for analysis of X-ray CT images in preclinical studies of mice for polycystic kidney disease. *IEEE Trans. Med. Imag.*, 21(10):1302–1309, 2002. ix, 19, 20
- [7] S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995. x, xiii, 9, 29, 30, 31, 57, 58
- [8] B. van Ginneken, A.F. Frangi, J.J. Staal, B.M. ter Haar Romeny, and M.A. Viergever. Active shape model segmentation with optimal features. *IEEE Trans. Med. Imag.*, 21(8):924–933, 2002. xii, 2, 12, 13, 14, 15, 40, 67, 69, 80, 84, 85
- [9] J. J. Koenderink and A.J. Vandoorn. The structure of locally orderless images. *Int. J. Comput. Vis.*, 31(2):159–168, 1999. xii, 15, 40
- [10] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957. xiii, 13, 18, 51
- [11] B. van Ginneken, M. B. Stegmann, and M. Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis*, 10(1):19–40, 2006. xiv, 9, 10, 12, 17, 27, 67, 94, 96, 97

- [12] SCR database: Segmentation in Chest Radiographs
<http://www.isi.uu.nl/Research/Databases/SCR/>. xiv, 27, 94, 96
- [13] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174:71–74, 2000. xiv, 94
- [14] Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Trans. Med. Imag.*, 20(7):595–604, 2001. 2
- [15] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E.A. Hoffman. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE Trans. Med. Imag.*, 25(4):464–475, 2006. 2
- [16] T. Ishida, S. Katsuragawa, K. Chida, H. MacMahon, and K. Doi. Computer-aided diagnosis for detection of cardiomegaly in digital chest radiographs. In *Proc. SPIE Conference on Medical Imaging*, volume 5747, San Diego, US, February 2005. 2
- [17] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Med. Imag.*, 20(8):677–688, 2001. 2
- [18] H.P. Meinzer, M. Thorn, and C.E. Cardenas. Computerized planning of liver surgery - an overview. *Computers and Graphics*, 26:569–576, 2002. 2, 104
- [19] P.P. Smyth, C.J. Taylor, and J.E. Adams. Automatic measurement of vertebral shape using active shape models. *Image and Vision Computing*, 15(8):575–581, 1997. 2, 12
- [20] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens. Evaluation of image features and search strategies for segmentation of bone structures using active shape models. *Medical Image Analysis*, 6(1):47–62, 2002. 2, 12, 20, 53
- [21] N. Duta and M. Sonka. Segmentation and interpretation of MR brain images: An improved active shape model. *IEEE Trans. Med. Imag.*, 17(6):1049–1067, June 1998. 2, 12
- [22] H.C. van Assen, M.G. Danilouchkine, F. Behloul, H.J. Lamb, R.J. van der Geest, J.H.C. Reiber, and B.P.F. Lelieveldt. Cardiac LV segmentation using a 3D active shape model driven by fuzzy inference. In *Medical Image Computing and Computer-Assisted Intervention*, volume 2878 of *Lecture Notes in Computer Science*, pages 533–540, Montreal, Canada, 2003. 2, 20
- [23] M.A. Fischler and R.A. Elschlager. The representation and matching of pictorial structures. *IEEE Trans. Comput.*, 22(1):67–92, 1973. 3, 12, 20
- [24] Y. Amit and A. Kong. Graphical templates for model registration. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(3):225–236, 1996. 13, 53

- [25] J. Coughlan, A. Yuille, C. English, and D. Snow. Efficient deformable template detection and localization without user initialization. *Computer Vision and Image Understanding*, 78(3):303–319, 2000. 13, 53
- [26] P.F. Felzenszwalb and D.P. Huttonlocker. Efficient matching of pictorial structures. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 66–73, South Carolina, US, June 2000. 3, 13, 17, 20
- [27] J.A. Hartigan and M.A. Wong. A K-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979. 7
- [28] F.J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 8:679–698, 1986. 7
- [29] U. Montanari. On the optimal detection of curves in noisy pictures. *Communications of the ACM*, 14:335–345, 1971. 7, 53
- [30] A. Martelli. An application of heuristic search methods to edge and contour detection. *Communications of the ACM*, 19(2):73–83, 1976. 7
- [31] P. Felzenszwalb and D. P. Huttonlocker. Efficient graph-based image segmentation. *Int. J. Comput. Vis.*, 59(2):167–181, 2004. 7
- [32] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 15(11):1101–1113, 1993. 7
- [33] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):888–905, 2000. 7
- [34] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Proc. IEEE International Conference on Computer Vision*, volume 1, pages 26–33, Nice, France, October 2003. 7
- [35] K. Li, X. Wu, D. Z. Chen, and M. Sonka. Optimal surface segmentation in volumetric images - a graph-theoretic approach. *IEEE Trans. Pattern Anal. Machine Intell.*, 28(1):119–134, 2006. 7
- [36] B. van Ginneken and B.M. ter Haar Romeny. Automatic segmentation of lung fields in chest radiographs. *Medical Physics*, 27(10):2445–2455, 2000. 8, 9, 67
- [37] E.N. Mortensen and W.A. Barret. Intelligent scissors for image composition. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191 – 198, Vancouver, Canada, 1995. 8
- [38] M.F. McNitt-Gray, H.K. Huang, and J.W. Sayre. Feature selection in the pattern classification problem of digital chest radiograph segmentation. *IEEE Trans. Med. Imag.*, 14(3):537–547, 1995. 9, 67
- [39] O. Tsujii, M.T. Freedman, and S.K. Mun. Automated segmentation of anatomical regions in chest radiographs using an adaptive-sized hybrid neural network. *Medical Physics*, 25(6):998–1007, 1998. 9

- [40] M. Loog and B. van Ginneken. Supervised segmentation by iterated contextual pixel classification. In *Proc. IEEE International Conference on Pattern Recognition*, volume 2, pages 925–928, Quebec, Canada, August 2002. 9, 10
- [41] N.F. Vittitoe, R. Vargas-Voracek, and C.E. Floyd Jr. Identification of lung regions in chest radiographs using Markov Random Field modeling. *Medical Physics*, 25(6):976–985, 1998. 9, 67
- [42] K. Van Leemput, F. Meas, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.*, 18(10):897–908, 1999. 9, 10, 19
- [43] Q. Wang, E. D’Agostino, D. Seghers, F. Maes, D. Vandermeulen, and P. Suetens. Large-scale validation of non-rigid registration algorithms for atlas-based brain image segmentation. In *Proc. SPIE Conference on Medical Imaging*, volume 6144, San Diego, US, February 2006. 10
- [44] T. Rohlfing, R. Brandt, R. Menzel, and J.C.R. Maurer. Segmentation of three-dimensional images using non-rigid registration: Methods and validation with application to confocal microscopy images of bee brains. In *Proc. SPIE Conference on Medical Imaging*, volume 5032, pages 363–374, San Diego, US, February 2003.
- [45] E. van Rikxoort, Y. Arzhaeva, and B. van Ginneken. A multi-atlas approach to automatic segmentation of the caudate nucleus in MR brain images. In T. Heimann, M. Styner, and B. van Ginneken, editors, *3D Segmentation in The Clinic: A Grand Challenge*, pages 29–36, Brisbane, Australia, October 2007. 10
- [46] A. Hammers, M.J. Koep, S.L. Free, M. Brett, M.P. Richardson, C. Labb, V.J. Cunningham, D.J. Brooks, and J. Duncan. Implementation and application of a brain template for multiple volumes of interest. *Human Brain Mapping*, 15(3):165–174, 2002. 10
- [47] D. Seghers, E. D’Agostino, F. Maes, D. Vandermeulen, and P. Suetens. Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques. In *Medical Image Computing and Computer-Assisted Intervention*, volume 3216 of *Lecture Notes in Computer Science*, pages 696–703, Saint-Malo, France, September 2004.
- [48] Q. Wang, D. Seghers, E. D’Agostino, F. Maes, D. Vandermeulen, P. Suetens, and A. Hammers. Construction and validation of mean shape atlas templates for atlas-based brain image segmentation. In *Information Processing in Medical Imaging*, volume 4057 of *Lecture Notes in Computer Science*, pages 136–143, Colorado, US, July 2005.
- [49] T. Rohlfing, R. Brandt, R. Menzel, and J.C.R. Maurer. *The Handbook of Medical Image Analysis: Segmentation and Registration Models*, chapter Quo Vadis, Atlas-Based Segmentation?, pages 435–486. Kluwer, 2005. 10
- [50] A.M. Ali, A.A. Farag, and A.S. El-Baz. Graph cuts framework for kidney segmentation with prior shape constraints. In *Medical Image Computing and Computer-Assisted Intervention*, volume 4791 of *Lecture Notes in Computer Science*, pages 384–392, Brisbane, Australia, October 2007. 10

- [51] A. L. Yuille, D. S. Cohen, and P. W. Hallinan. Feature extraction from faces using deformable templates. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pages 104–109, San Diego, US, June 1989. 11
- [52] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour model. *Int. J. Comput. Vis.*, 1(4):321–333, 1988. 11
- [53] J. A. Sethian. *Level set methods and fast marching methods*. Cambridge University Press, 1999. 11
- [54] T.F. Cootes and C.J. Taylor. Statistical models of appearance for computer vision. Technical report, Wolfson Image Anal. Unit, Univ. Manchester, Manchester, U.K., 2004. 12, 13, 14, 95
- [55] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and applications. *Computer Vision and Image Understanding*, 61(1):38–59, 1995. 12, 85
- [56] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(6):681–685, 2001. 12, 16, 17
- [57] T.F. Cootes and C.J. Taylor. Constrained active appearance models. In *Proc. IEEE International Conference on Computer Vision*, volume 1, pages 748–754, Vancouver, Canada, July 2001.
- [58] M.B. Stegmann, B.K. Ersbøll, and R. Larsen. Extending and applying active appearance models for automated and high precision segmentation in different image modalities. In *Proc. 12th Scandinavian Conference on Image Analysis*, pages 90–97, Bergen, Norway, June 2001. 12, 17
- [59] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Comparing active shape models with active appearance models. In *Proc. British Machine Vision Conference*, pages 173–182, Nottingham, UK, September 1999. 12
- [60] R. Beichel, H. Bischof, F. Leberl, and M. Sonka. Robust active appearance models and their application to medical image analysis. *IEEE Trans. Med. Imag.*, 24(9):5–69, 2005. 12, 17
- [61] S. Mitchell, B. Lelieveldt, R. van der Geest, H. Bosch, J. Reiber, and M. Sonka. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. *IEEE Trans. Med. Imag.*, 20(5):415–423, 2001.
- [62] E. Oost, G. Koning, M. Sonka, P.V. Oemrawsingh, J.H. Reiber, and B.P. Lelieveldt. Automated contour detection in X-ray left ventricular angiograms using multiview active appearance models and dynamic programming. *IEEE Trans. Med. Imag.*, 25(9):1158–1171, 2006. 12
- [63] B. van Ginneken, S. Katsuragawa, B.M. ter Haar Romeny, K. Doi, and M.A. Viergever. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Trans. Med. Imag.*, 25(6):998–1007, 2002. 12

- [64] T. Heimann, I. Wolf, and H.P. Meinzer. Active shape models for a fully automated 3D segmentation of the liver - An evaluation on clinical data. In *Medical Image Computing and Computer-Assisted Intervention*, volume 4791 of *Lecture Notes in Computer Science*, pages 86–93, Brisbane, Australia, October 2006. 12, 104
- [65] Y. Wang and L. Staib. Boundary finding with correspondence using statistical shape models. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 338–345, Santa Barbara, US, June 1998. 12, 20
- [66] M. Rogers and J. Graham. Robust active shape model search for medical image analysis. In *Proc. European Conference on Computer Vision*, pages 517–530, Copenhagen, Denmark, May 2002.
- [67] L. Shuyu, Z. Litao, and J. Tianzi. Active shape model segmentation using local edge structures and AdaBoost. In *Second International Workshop, Medical Imaging and Augmented Reality*, volume 3150 of *Lecture Notes in Computer Science*, pages 121–128, Beijing, China, August 2004. 12
- [68] P.F. Felzenszwalb and D.P. Huttonlocker. Pictorial structures for object recognition. *Int. J. Comput. Vis.*, 61(1):55–79, 2005. 13, 20
- [69] D. Cristinacce and T. Cootes. Facial feature detection and tracking with automatic template selection. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 429–434, Southampton, UK, April 2006. 13
- [70] C. Goodall. Procrustes methods in the statistical analysis of shapes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339, 1991. 14
- [71] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Temporal subtraction of thorax CR images using a statistical deformation model. *IEEE Trans. Med. Imag.*, 22(11):1490–1504, 2003. 14
- [72] G.J. Edwards, T.F. Cootes, and C.J. Taylor. Advances in active appearance models. In *Proc. IEEE International Conference on Computer Vision*, volume 1, pages 137–142, Corfu, Greece, September 1999. 17
- [73] P.F. Felzenszwalb. Representation and detection of deformable shapes. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(2):208–220, 2005. 17, 53
- [74] Y. Tianly, L. Jiebo, S. Amit, and A. Narendra. Shape regularized active contour based on dynamic programming for anatomical structure segmentation. In *Proc. SPIE Conference on Medical Imaging*, pages 419–430, San Diego, US, February 2005. 20, 53, 67
- [75] S. Gleason, H. Sari-Sarraf, M. Paulus, D. Johnson, and M. Abidi. Automatic screening of polycystic kidney disease in X-ray CT images of laboratory mice. In *Proc. SPIE Conference on Medical Imaging*, volume 3979, pages 837–846, San Diego, US, February 2000. 20

- [76] J. Weese, M. Kaus, C. Lorenz, S. Lobregt, R. Truyen, and V. Pekar. Shape constrained deformable models for 3D medical image segmentation. In *Information Processing in Medical Imaging*, volume 2082 of *Lecture Notes in Computer Science*, pages 380–387, Davis, US, June 2001. 20
- [77] D. Kainmiller, T. Lange, and H. Lamecker. Shape constrained automatic segmentation of the liver based on a heuristic intensity model. In T. Heimann, M. Styner, and B. van Ginneken, editors, *3D Segmentation in The Clinic: A Grand Challenge*, pages 109–116, Brisbane, Australia, October 2007. 20, 104
- [78] Segmentation of the liver 2007 (SLIVER07). <http://sliver07.org/>. 27, 104, 105, 106
- [79] P.J. Besl and N.D. McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(2):239–256, 1992. 28
- [80] P. Claes. *A robust statistical surface registration framework using implicit function representations - Application in craniofacial reconstruction*. PhD thesis, Katholieke Universiteit Leuven, 2007. 28, 107, 108
- [81] R. H. Davies, C.J. Twining, T.F. Cootes, J.C. Waterton, and C.J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Trans. Med. Imag.*, 21(5):525–537, 2002. 28
- [82] M. A. Styner, K.T. Rajamani, L.P. Nolte, G. Zsemlye, G. Szekely, C. J. Taylor, and R.H. Davies. Evaluation of 3D correspondence methods for model building. In *Information Processing in Medical Imaging*, volume 2732 of *Lecture Notes in Computer Science*, pages 63–75, Ambleside, UK, July 2003. 28
- [83] T. Heimann, I. Wold, T. Williams, and H.P. Meinzer. 3D active shape models using gradient descent optimization of description length. In *Information Processing in Medical Imaging*, volume 3565 of *Lecture Notes in Computer Science*, pages 566–577, Glenwood Springs, US, July 2005. 28
- [84] T.L. Tong. *The Multivariate Normal Distribution*. Springer-Verlag, 1990. 36, 121
- [85] Wikipedia: Maximum likelihood. <http://en.wikipedia.org/>. 36
- [86] B. van Ginneken and B. M. H. ter Haar Romeny. Applications of locally orderless images. In *Proceedings International Conference on Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes In Computer Science*, pages 10–21, Corfu, Greece, September 1999. 41
- [87] B. van Ginneken, T. Heimann, and M. Styner. 3D segmentation in the clinic: A grand challenge. In T. Heimann, M. Styner, and B. van Ginneken, editors, *3D Segmentation in The Clinic: A Grand Challenge*, pages 7–15, Brisbane, Australia, October 2007. 104, 109
- [88] P. Slagmolen, A. Elen, D. Seghers, D. Loeckx, F. Maes, and K. Haustermans. Atlas based liver segmentation using nonrigid registration with a B-spline transformation model. In T. Heimann, M. Styner, and B. van Ginneken, editors, *3D Segmentation in The Clinic: A Grand Challenge*, pages 197–206, Brisbane, Australia, October 2007. 106

- [89] Matlab FastRBF Interpolation Toolbox .
<http://www.farfieldtechnology.com/products/toolbox>. 107
- [90] N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: message-passing revisited. In *Proc. IEEE International Conference on Computer Vision*, pages 1–8, Rio de Janeiro, Brazil, October 2007. 118
- [91] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Anal. Machine Intell.*, 29(7):1165–1179, 2007. 118
- [92] E. D’Agostino. *An information theoretic framework for atlas-based brain image registration and segmentation*. PhD thesis, Katholieke Universiteit Leuven, 2006. 119

Publications

International Journal

1. D. Seghers, D. Loeckx, F. Maes, D. Vandermeulen and P. Suetens. Minimal Intensity and Shape Cost Path Segmentation *IEEE Transactions on Medical Imaging*, vol. 26, no. 8, pp. 1115-1129, August 2007
2. T. Heimann and B. van Ginneken and M. Styner and Y. Arzhaeva and A. Volker and C. Bauer and A. Beck and C. Becker and R. Beichel and G. Bekes and F. Bello and G. Binnig and H. Bischof and A. Bornik and P.M.M. Cashman Y. Chi and A. Córdova and B.M. Dawant and M. Fidrich and J. Furst and D. Furukawa and L. Grenacher and J. Hornegger and D. Kainmüller and R. I. Kitney and H. Hobatake and H. Lamecker and T. Lange and J. Lee and B. Lennon and R. Li. S. Li and H.P. Meinzer and G. Németh and D.S. Raicu and A.M. Rau and E. van Rikxvoort and M. Rousson and L. Ruskó and K.A. Saddi and G. Schmidt and D. Seghers and A. Shimizu and P. Slagmolen and E. Soranting and G. Soza and R. Susomboon and J.M. Waite and A. Wimmer and I. Wolf. Comparison and evaluation of methods for liver segmentation from CT datasets *IEEE Transactions on Medical Imaging*, 2008 (submitted)

International Conference : Published in Proceedings

1. D. Seghers, E. D'Agostino, F. Maes, D. Vandermeulen and P. Suetens. Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques *Lecture notes in computer science*, vol. 3216, pp. 696-703, 2004 (Proceedings 7th international conference on medical image computing and computer-assisted intervention - MICCAI 2004, September 26-29, 2004, Saint-Malo, France)
2. Q. Wang, D. Seghers, E. D'Agostino, F. Maes, D. Vandermeulen, P. Suetens and A. Hammers. Construction and validation of mean shape atlas templates for atlas-based brain image segmentation *Lecture notes*

- in computer science*, vol. 3565, pp. 689-700, 2005, Springer-Verlag (Proceedings information processing in medical imaging - IPMI 2005, July 11-15, 2005, Glenwood Springs, Colorado, USA)
3. S. Drisis, D. Seghers, S. Srivastava, E. D'Agostino, F. Maes and G. Marchal. Non-rigid registration for subtraction CT angiography of the abdomen and evaluation of various regularization models *17th European congress of radiology - ECR 2005*, paper no. B-551, March 4-8, 2005, Vienna, Austria
 4. S. Drisis, D. Seghers, S. Srivastava, E. D'Agostino, F. Maes and G. Marchal. Non-rigid image registration for subtraction CT angiography of the abdomen is significantly more effective when the post-contrast image is deformed to match the pre-contrast image *17th European congress of radiology - ECR 2005*, poster no. C-1045, March 4-8, 2005, Vienna, Austria
 5. S. Drisis, S. Srivastava, D. Seghers, W. Coudyzer, E. D'Agostino, F. Maes, P. Suetens and G. Marchal. Subtraction CT angiography using non-rigid registration : the impact of similarity measure and image pre-processing *Proceedings 19th international congress and exhibition on computer assisted radiology and surgery - CARS 2005*, pp. 328-333, June 22-25, 2005, Berlin, Germany
 6. D. Seghers, D. Loeckx, F. Maes and P. Suetens. Image segmentation using local shape and gray-level appearance models *Proceedings SPIE medical imaging 2006 conference*, vol. 6144, 10 pp., February 11-16, 2006, San Diego, California, USA
 7. Q. Wang, E. D'Agostino, D. Seghers, F. Maes, D. Vandermeulen and P. Suetens. Large-scale validation of non-rigid registration algorithms for atlas-based brain image segmentation *Proceedings SPIE medical imaging 2006 conference*, vol. 6144, 8 pp., February 11-16, 2006, San Diego, California, USA
 8. D. Seghers, D. Loeckx, F. Maes and P. Suetens. Visual enhancement of interval changes using a temporal subtraction technique *SPIE medical imaging 2007 conference*, vol. 6512, 9 pp., February 17-22, 2007, San Diego, California, USA (honorable mention poster award)
 9. D. Seghers, P. Slagmolen, Y. Lambelin, J. Hermans, D. Loeckx, F. Maes and P. Suetens. Landmark based liver segmentation using local shape and local intensity models *Workshop proceedings of the 10th international conference on medical image computing and computer assisted intervention - MICCAI 2007, workshop on 3D segmentation in the clinic : a grand challenge*, pp. 135-142, October 29, 2007, Brisbane, Australia
 10. P. Slagmolen, A. Elen, D. Seghers, D. Loeckx, F. Maes, K. Haustermans. Atlas based liver segmentation using nonrigid registration with a

B-spline transformation model *Workshop proceedings of the 10th international conference on medical image computing and computer assisted intervention - MICCAI 2007, workshop on 3D segmentation in the clinic : a grand challenge*, pp. 197-206, October 29, 2007, Brisbane, Australia

11. D. Seghers, J. Hermans, D. Loeckx, F. Meas, D. Vandermeulen and P. Suetens. Model based segmentation using graph representations *Lecture notes in computer science*, vol. 5241, pp. 393-400, 2008 (Proceedings 11th international conference on medical image computing and computer-assisted intervention - MICCAI 2008, September 6-10, 2008, New York, USA)

International Conference : Abstract or Not Published

1. P. Slagmolen, A. Elen, D. Seghers, D. Loeckx, F. Maes and K. Haustermans. Atlas based liver segmentation using nonrigid registration *Proceedings Annual symposium of the IEEE/EMBS Benelux chapter*, December 6-7, 2007, Heeze, The Netherlands

Internal Report

1. D. Seghers. Design of a MRI brain atlas *Internal report KUL/ESAT/PSI/0304*, K.U.Leuven, ESAT, December 2003, Leuven, Belgium
2. Q. Wang, E. D'Agostino, D. Seghers, F. Maes, D. Vandermeulen and P. Suetens. Large-scale validation of non-rigid registration algorithms in a brain image segmentation framework *Internal report KUL/ESAT/PSI/0502*, K.U.Leuven, ESAT, March 2005, Leuven, Belgium
3. Q. Wang, E. D'Agostino, D. Seghers, F. Maes, D. Vandermeulen and P. Suetens. Construction and validation of statistical brain atlases for atlas-based brain image segmentation *Internal report KUL/ESAT/PSI/0503*, K.U.Leuven, ESAT, March 2005, Leuven, Belgium

Patent

1. P. Dewaele, D. Loeckx and D. Seghers. Method of constructing a gray value model and/or a geometric model of an anatomic entity in a digital medical image *European Patent Office*, EP 5 107 903.3, August 30, 2005 (submitted)
2. P. Dewaele, D. Loeckx and D. Seghers. Method of segmenting anatomic entities in digital medical images *European Patent Office*, EP 5 107 907.7, August 30, 2005 (submitted)

