

V4Design for enhancing architecture and video game creation

Konstantinos Avgerinakis^{*1}, George Meditskos¹, Jens Derdaele², Simon Mille³, Yash Shekhawat⁴, Luis Fraguada⁵, Eva Lopez⁶, Jan Wuyts⁷, Anastasios Tellios⁸, Steffen Riegas⁹, Jesper Wachtmeister¹⁰, kriszta Doczy¹¹, Victor-Jan Vos⁷, Nicolaus Heise⁶, Jens Piesk⁴, Maarten Vergauwen², Leo Wanner³, Stefanos Vrochidis¹, and Ioannis Kompatsiaris¹

¹Centre for Research and Technology Hellas, CERTH-ITI

²Katholieke Universiteit Leuven, KUL

³Universitat Pompeu Fabra, UPF

⁴Nurogames, NURO

⁵McNeel, McNeel Europe SL

⁶Deutsche Welle, DW

⁷Europeana Foundation, EF

⁸Aristotle University of Thessaloniki, AUTH

⁹Herzog & de Meuron, HdM

¹⁰Solaris Film produktion, SFp

¹¹ArtFilms, AF

ABSTRACT

Architecture, Design and Virtual Reality game creation companies are in great need of acquiring, re-using and re-purposing visual and textual data to recreate, renovate or produce a novel target space, building or element. This comes in alignment with the abrupt increase, which is lately observed, in the use of immersive VR environments and the great technological advance that can be found in the acquisition and manipulation of digital models. V4Design is a new project that takes into account these needs and intends to build a complete and robust design solution that will help the targeted industries to improve their creation process.

Index Terms: Computer vision—3D-reconstruction—VR in architecture—VR video gametext summarization;

1 INTRODUCTION

Vast troves of visual and textual data, which are of great interest to architects and video game designers, such as paintings, archival footage, documentaries, movies, reviews or catalogues, and various other forms of artwork, are currently difficult to access and cannot be easily incorporated in the design process. Such works of art can serve as sources of inspiration and design direction towards innovative designs, new concepts, or period-focused structures, among others.

V4Design project aims to enable the **re-use** and **re-purpose** of such multimedia content by proposing an innovative solution for its effective integration, so as to extract 3D representations and VR game environments in a semi-automatic way. This will allow architects, designers and video game creators to re-use heterogeneous archives of already available and retrieved digital content and re-purpose it by making the wealth of 3D, VR, aesthetic and textual information easily accessible and providing resources and tools to design and model outdoors and indoors environments of architecture and VR video game projects.

To achieve this, V4Design builds upon the concept of semantic integration of heterogeneous 2-dimensional multimedia in order to generate enhanced dynamic 3-dimensional structures and environments. Towards this direction V4Design collects data from the consortium partners (DW, EF, AF, SFp) and by crawling online

*e-mail: koafgeri@iti.gr

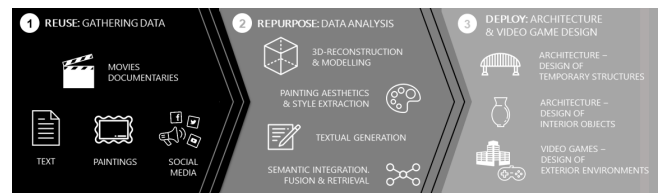


Figure 1: Abstract V4Design tool architecture

art libraries and databases for retrieving more visual and textual content. The compiled content is mainly used to create dynamic 3D-models of objects, buildings/structures and environments and use them to support the creation of architecture and video game VR environments. Architects and video game designers are thus able to easily access and use a realistic and comprehensive representation of structures of interest from various time periods and styles in an economical and cost-effective manner. Innovative results that integrate past aesthetic trends, localization of buildings and art-elements in visual data, text generation and semantic knowledge are also produced as complementary material to support the design process.

The ultimate goal of V4Design platform is to enable the efficient re-purposing and re-use of existing digital content, extending its lifespan and maximizing the return on investment for producers. The corresponding market (e.g. movie or documentary production) will expand into new domains, those of architecture and video game design as the V4Design content production and management tools will provide efficient access to a wealth of previously unused data.

2 V4DESIGN INDIVIDUAL TOOLS

V4Design utilizes a great deal of architecture and video game design innovative tools to enhance and automate the re-use and re-purpose of textual and visual digital content. Namely, V4Design incorporates (i) data collection and web crawling; (ii) spatio-temporal localization of objects and buildings in visual data; (iii) 3D-reconstruction for 3D-model extraction; (iv) aesthetics extraction and style transfer from paintings and art imagery; (v) content extraction from textual data; (vi) semantic representation, integration and reasoning of all the above and finally (vii) the implementation of innovative architecture, design and VR game authoring applications that will incorporate and show all the appropriate information to V4Design users. An abstract scheme of V4Design platform is depicted in Fig. 1.

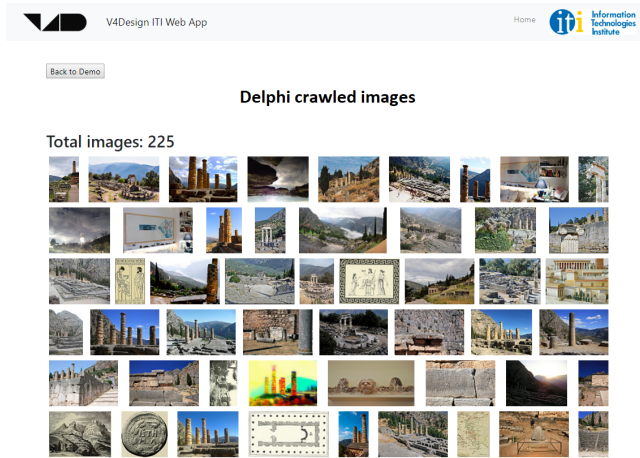


Figure 2: Crawling data for Delphi showcase

2.1 Data collection and web crawling

Data collection and web crawling component revolves around gathering of the required content (i.e. architecture visual data from movies, documentaries, paintings, other artwork pictures, and relevant textual content) to build a complete and meaningful database for running V4Design tools.

The project relies upon existing data-sharing services and develop data wrappers in order to support content access to the architecture and art-related film producers and distributors in the consortium: ArtFimls (AF), Solaris Filmproduktion (SFP) and Deutsche Welle (DW), who will provide selected parts of their documentary and movie archives to V4Design platform so as to localize building structures and objects and build their 3D models. Europeana (EF) will also provide their large archive of paintings, pictures of contemporary artwork and related critics, for stylistic and aesthetics extraction and textual analysis.

Visual artwork is also crawled from public databases, by deploying advanced retrieval techniques [19] [12] that collect paintings, images, as well as related textual content, such as critics, catalogues and museum guides, retrieved from social media, movie-related free archives and other open sources. Crawling data for Delphi showcase are depicted in Fig. 2.

2.2 Spatio-Temporal Building and Object Localization

Spatio-Temporal Building and Object Localization (STBOL) is applied on art and architecture-related movies, documentaries and multiple art-images, aiming to localize and semantically segment the interiors of buildings, including objects in them, as well as building exteriors and landscapes. For the purposes of interior architecture design, objects will be automatically detected, recognized and localized inside video frames, using State-of-the-Art (SotA) Deep Convolutional Neural Network (DCNN) object detection techniques, such as Faster RCNN [13] and SSD [8]. For building exteriors, scene recognition algorithms that are based on PlacesCNN [21] have been deployed and applied in history and modern city architecture movies and documentaries so as to localize modern and historic cityscape environments. As far as spatial localization is concerned, semantic segmentation algorithms, inspired from the recent success of DeepLab [3], have been deployed in architecture environments and are provided in 3D-reconstruction component so as to tag the extracted 3D-models appropriately. Fig. 3 depict a characteristic scene from Delphi showcase and the retrieved image tags.



Figure 3: Semantic segmentation of Delphi temple. Retrieved tags separate semantically the image in the following regions: sky, mountain, columns and ancient ruins

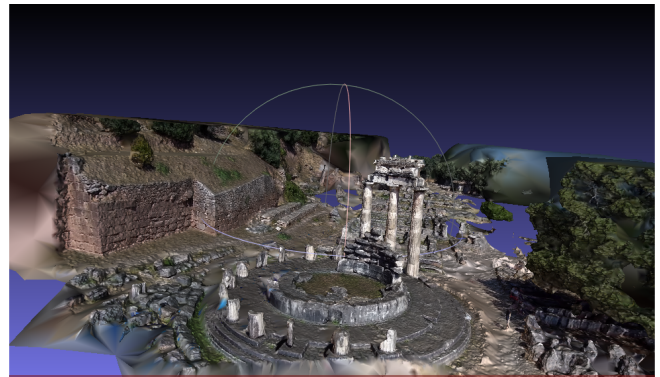


Figure 4: Delphi temple 3D model

2.3 3D-reconstruction for 3D-model extraction

3D-reconstruction is a fundamental component in V4Design platform and is responsible for the extraction of 3D-model structures out of the retrieved video sequences. As a pre-processing step, 3D-reconstruction analyzes the input sequences and assess their suitability for extracting a robust and complete 3D-model. This is performed by using a combination of keyframe selection, GRIC scores [18], segmentation algorithms and fast Structure-from-Motion (SfM) procedures. The most promising sequences are processed further and more thorough camera calibration and sparse point clouds are computed. Self-calibration algorithms are employed if no prior information about the camera is available. The sparse point clouds are cleaned in a (semi-)automatic way and dense 3D point clouds and textured meshes are computed, using SoA algorithms like semi-global matching [7] or PMVS [20]. Further investigation is to be made in this field by studying recent DCNN techniques that use crawled data to retrieve the 3D model of a specific region [15]. Characteristic examples of this procedure are depicted in Fig. 4.

2.4 Aesthetics extraction and texture proposals

The Aesthetic Extraction (AE) from paintings component aims to extract and categorize the aesthetics of paintings based on their style (i.e. impressionism, cubism and expressionism), creators and emotion that they evoke to the viewer. This component is based on a DCNN, trained in 80K images of the WikiArt dataset [14], and is



Figure 5: Aesthetic style transfer to Delphi showcase

used to categorize the acquired paintings so as to create an aesthetic gallery from where an end-user could choose or be inspired from in order to create novel architecture structures or other artworks.

Furthermore, Texture Proposals (TP) from paintings and other artwork images use the categorized paintings, extracted from AE component, so as to create novel aesthetics and style textures which will be provided to architects and video game designers so that they can create novel artwork and structures based on past observations and styles. This technology is based on SotA style transfer algorithms that have recently proposed in DCNN literature [5, 6]. Characteristic aesthetics style transfer example of this technique can be found in Fig. 5.

2.5 Extracting content from textual data

Content extraction from textual data is in charge of generating textual reports, descriptions, or summaries, starting from data extracted from text, webpages, and/or visual analytics. The component initially builds abstract representations, modeled as RDF triples, which are stored in a semantic repository, and then continues with a request for a summary of most relevant contents related to a specific keyword (or entity) or comes along a generated 3D model.

The Language Analysis module follows the above procedure and continues with the analysis and capturing of the natural language textual material into structured, ontological representations, so that appropriate system responses can subsequently be inferred by the reasoning module (CERTH), and that textual summaries can be produced (TALN-Language Generation). The module combines multilingual dependency parsers and lexical resources, and a projection of the extracted dependency-based linguistic representations into ontological ones. Our current algorithms in textual analysis are based on Babelnet [11] and DBpedia [1] and their results in Delphi showcase can be seen in Fig. 6.

2.6 Semantic content representation, integration and reasoning

Semantic content representation, integration and reasoning focuses on the knowledge data integration of all the system components into an appropriately ontology-modelled network to support unified retrieval. Semantic reasoning follows so as to further enrich this information by deriving implicit information from the hierarchies of concepts and from their interrelationships. More specifically, content modelling is responsible for delivering a semantic representation of the textual and visual derived concepts, such as BIM and GIS models and natural language conceptual representations, in the form of a network of interconnected ontologies. Semantic integration is based on this representation so as to semantically enrich the ontology network with information from appropriate external sources

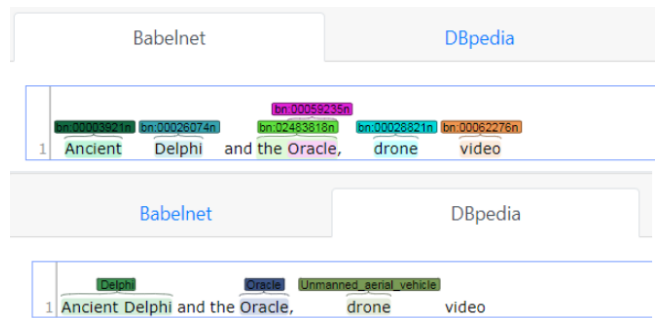


Figure 6: Text analysis deployed in Delphi showcase

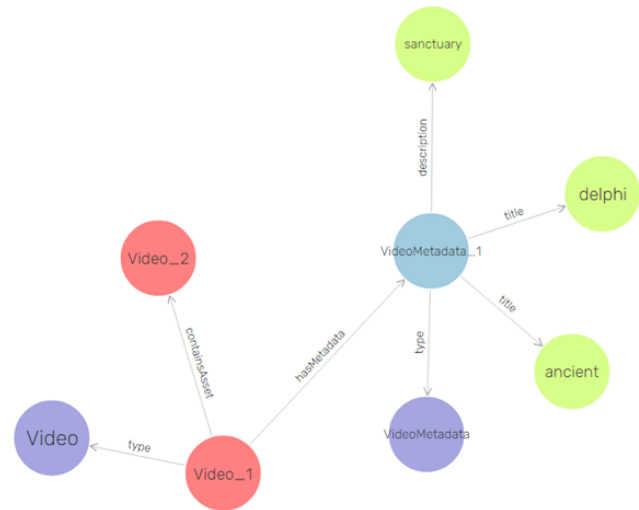


Figure 7: Semantic content representation of Delphi showcase

(structured, semi-structured, unstructured) and semantic reasoning follows up in order to support the discovery of implicit relationships between concepts in the ontologies, allowing for enhanced indexing and retrieval capabilities of the system.

The formal representation of the developed ontological models comply with the W3C established ontology language OWL2, while for the development of the ontologies, V4Design relies on popular ontology engineering methodologies, such as NeOn [17], aiming at modular but at the same time manageable and maintainable semantic models. Existing core models are re-used and extended and popular sources will be consulted for appropriate third-party models, like e.g. the Linked Open Data Vocabularies (LOV) or Ontology Design Patterns (ODP). Established RDF triple stores will be used for hosting the models. Semantic integration/enrichment relies on high-end research approaches, such as PROPheT ontology populator [9] ontology alignment tools from the MULTISENSOR project [4], DL reasoners, like Pellet [16] and Hermit [10]. While, uncertainty reasoners for handling uncertain information will also be investigated using advanced techniques, such as FuzzyDL [2]. Current work focus on the construction of the ontology representation and a sample of Delphi showcase can be seen in Fig. 7. There we can see how text, videos and metadata derived from the abovementioned components are linked to each other.

3 ARCHITECTURE AND VIDEO GAME DESIGN SHOWCASES

This section will elaborate on the two showcases that V4Design plans to invest so as to demonstrate its capabilities.



Figure 8: Example of Delphi showcase

3.1 Architectural design of buildings, landscapes and smaller artistic elements

The pilot for Architectural design, related to existing or historical buildings and their Environments is designed to use the capabilities of V4Design so as to acquire digital 3D models of existing architecture, whether it is about landscapes, buildings and smaller scale artistic elements. This showcase uses existing visual and textual content from the material that is provided from the content providers (DW,EF,AF,SFp) and additionally free content available on the web.

The first showcase of V4Design for architectural design purposes is going to take place in the area of the archaeological site of Delphi in Greece Fig. 8. In this first showcase V4Design authoring tool is used so as to help the architects to reconstruct the landscape and the various spatial elements that articulate it. Further on, V4Design will support the designing of open-air infrastructures, such as pavilions, landart elements and scenographic arrangements for the archaeological site of Delphi.

V4design has designed a set of tools that are provided to architects so as to improve their design process. Data have been acquired from data providers and web crawling, as demonstrated in Fig. 2, and V4Design components have been deployed so as to offer aesthetic texture proposals (Fig. 5), 3D-models of the area (Fig. 4), textual information from relevant content (Fig. 6), as well as semantic segmentation in images of the region (Fig. 3). V4Design integrated (Fig. 7) and will import this information in Rhino authoring tool, as well as Virtual Reality (VR) immersive environment, so that the architects will be able to build their creation and design proposals.

3.2 Video game design based on archived footage

The showcase for re-living the date take places in Gendanmarkt square (Fig. 9) is designed in order to use the capabilities of V4Design for development of environments from archives of news and videos captured some years ago. The showcase depends on existing content from various content providers (i.e. DW), as well as look into new content that has been captured of a particular scene so as to create a similar environment as the real life throughout history. By this way, the user is able to interact with the environments and the same time gain more knowledge about the location or event.

In this case, V4design extracted the 3D models from various archived and more recent footage provided by DW and other content providers. The results were provided to video game creators (i.e. NURO), who used them so as to create Gendanmarkt square in a VR immersive environment and depict/highlight the changes of the square throughout time. Small scale(i.e. lamps) and large scale(i.e. buildings and sculptures) 3D-reconstructed models from the past can be seen in Fig. 10 and Fig. 11, Fig. 12 respectively.



Figure 9: Example of Gendanmarkt showcase

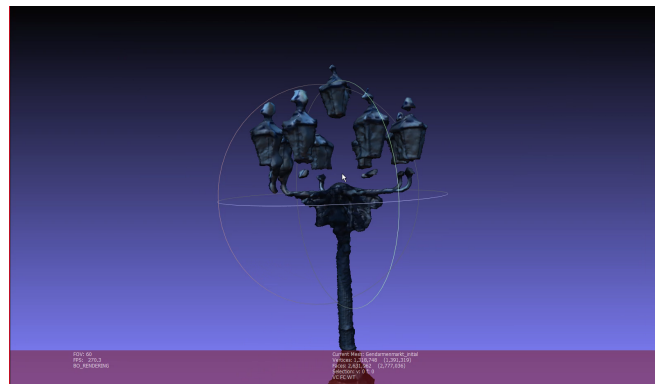


Figure 10: 3D-reconstructed model of a lamp in Gendanmarkt showcase

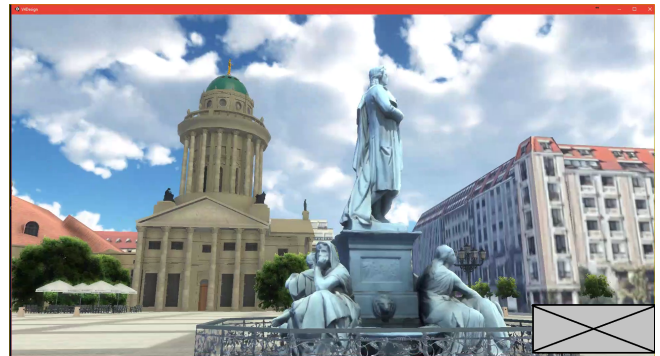


Figure 11: 3D-reconstructed model of the Schiller statue in Gendanmarkt showcase

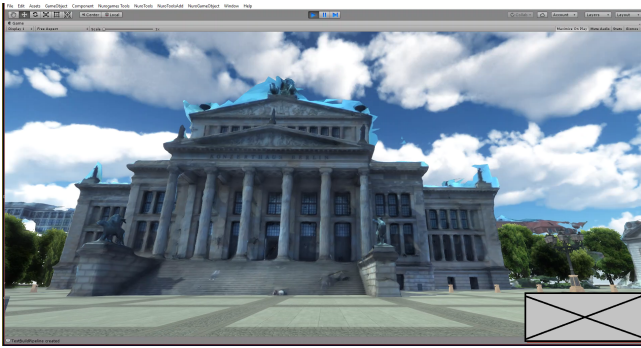


Figure 12: 3D-reconstructed model of the Konzerthaus in Gendarmenmarkt showcase

4 CONCLUSION AND FUTURE WORK

V4Design provides a user-friendly environment, where architects and video game and VR creators can effortlessly build, search and alter their creations. V4Design also intends in the near future to improve its data collection tools so that it can provide more and better visual data to 3D-reconstruction component, improve the algorithmic core of its components, such as the 3D-reconstruction, texture proposals, spatio-temporal building and object localization and textual analysis, so that it would provide a spate of complementary information based on the user needs. While, it also plans to provide a flexible and scalable integration and research system that would be easily to expanded in the future.

ACKNOWLEDGMENTS

This work was supported by V4Design project partially funded by the European Commission under grant agreements No 779962

REFERENCES

- [1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pp. 722–735. Springer, 2007.
- [2] F. Bobillo and U. Straccia. fuzzydl: An expressive fuzzy description logic reasoner. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, pp. 923–930. IEEE, 2008.
- [3] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.
- [4] C. Doulaverakis, S. Vrochidis, and I. Kompatsiaris. Exploiting visual similarities for ontology alignment. In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) KEOD; 2015 Nov 12-14; Lisbon, Portugal. Portugal: SCITEPRESS; 2015. vol. 2. p. 29-37*. SCITEPRESS, 2015.
- [5] L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman. Preserving color in neural artistic style transfer. *arXiv preprint arXiv:1606.05897*, 2016.
- [6] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423, 2016.
- [7] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 807–814. IEEE, 2005.
- [8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21–37. Springer, 2016.
- [9] P. Mitzias, M. Riga, E. Kontopoulos, T. G. Stavropoulos, S. Andreadis, G. Meditskos, and I. Kompatsiaris. User-driven ontology population

from linked data sources. In *International Conference on Knowledge Engineering and the Semantic Web*, pp. 31–41. Springer, 2016.

- [10] B. Motik, R. Shearer, and I. Horrocks. Hypertableau reasoning for description logics. *Journal of Artificial Intelligence Research*, 36:165–228, 2009.
- [11] R. Navigli and S. P. Ponzetto. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pp. 216–225. Association for Computational Linguistics, 2010.
- [12] S. Papadopoulos and Y. Kompatsiaris. Social multimedia crawling for mining and search. *IEEE Computer*, 47(5):84–87, 2014.
- [13] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99, 2015.
- [14] B. Saleh and A. Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*, 2015.
- [15] J. L. Schonberger, F. Radenovic, O. Chum, and J.-M. Frahm. From single image query to detailed 3d reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5126–5134, 2015.
- [16] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: science, services and agents on the World Wide Web*, 5(2):51–53, 2007.
- [17] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López. The neon methodology for ontology engineering. In *Ontology engineering in a networked world*, pp. 9–34. Springer, 2012.
- [18] P. H. Torr. An assessment of information criteria for motion model selection. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 47–52. IEEE, 1997.
- [19] T. Tsirikika, A. Latas, A. Moutzidou, E. Chatzilari, S. Vrochidis, and Y. Kompatsiaris. Discovery of environmental web resources based on the combination of multimedia evidence. In *Proceedings of the 2nd International Workshop on Environmental Multimedia Retrieval*, pp. 27–32. ACM, 2015.
- [20] X. Xiao, B. Guo, D. Li, L. Li, N. Yang, J. Liu, P. Zhang, and Z. Peng. Multi-view stereo matching based on self-adaptive patch and image grouping for multiple unmanned aerial vehicle imagery. *Remote Sensing*, 8(2):89, 2016.
- [21] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pp. 487–495, 2014.