

ONLINE 4D RECONSTRUCTION USING MULTI-IMAGES AVAILABLE UNDER OPEN ACCESS

Marinos Ioannides ^{a,*}, Andreas Hadjiprocopis ^a, Nikolaos Doulamis ^a, Anastasios Doulamis ^b, Eftychios Protopapadakis ^b, Kostas Makantasis ^b, Pedro Santos ^c, Dieter Fellner ^c, Andre Stork ^c, Olivier Balet ^d, Martine Julien ^d, Guenther Weinlinger ^e, Paul S.Johnson ^e, Michael Klein ^e, Dieter Fritsch ^f

^a Digital Heritage Lab, Cyprus University of Technology, Cyprus - marinos.ioannides@cut.ac.cy

^b Technical University of Crete, Kounopidiana, Chania, Greece – adoulam@ergasya.tuc.gr

^c Fraunhofer Institute for Computer Graphics Research IGD, Germany - pedro.santos@igd.fraunhofer.de

^d DigiNext, France - martine.julien@diginext.fr

^e 7Reasons GmbH, Austria - gw@7reasons.at

^f University of Stuttgart, Germany - dieter.fritsch@ifp.uni-stuttgart.de

KEY WORDS: Image Search, 3D Reconstruction, Computer Vision, Digital Libraries, Image Hosting Services, Visual Search, Augmentation

ABSTRACT:

The advent of technology in digital cameras and their incorporation into virtually any smart mobile device has led to an explosion of the number of photographs taken every day. Today, the number of images stored online and available freely has reached unprecedented levels. It is estimated that in 2011, there were over 100 billion photographs stored in just one of the major social media sites. This number is growing exponentially. Moreover, advances in the fields of Photogrammetry and Computer Vision have led to significant breakthroughs such as the Structure from Motion algorithm which creates 3D models of objects using their two-dimensional photographs. The existence of powerful and affordable computational machinery not only the reconstruction of complex structures but also entire cities. This paper illustrates an overview of our methodology for producing 3D models of Cultural Heritage structures such as monuments and artefacts from 2D data (pictures, video), available on Internet repositories, social media, Google Maps, Bing, etc. We also present new approaches to semantic enrichment of the end results and their subsequent export to Europeana, the European digital library, for integrated, interactive 3D visualisation within regular web browsers using WebGL and X3D. Our main goal is to enable historians, architects, archaeologists, urban planners and affiliated professionals to reconstruct views of historical structures from millions of images floating around the web and interact with them.

1. INTRODUCTION

The scope of economic activities of the Cultural and Creative industries focuses on the generation or exploitation of knowledge and information (Hesmondhalgh, 2007). According to (Florida, 2002) “human creativity is the ultimate economic resource,” and “the industries of the twenty-first century will depend increasingly on the generation of knowledge through creativity and innovation”. In Europe the rapid roll-out of new technologies and increased globalisation has meant a striking shift away from traditional manufacturing towards services and innovation. In Europe factory floors are progressively being replaced by creative communities whose raw material is their ability to imagine, create and innovate. Therefore, a major factor in maintaining European competitiveness is to unlock the potential of cultural and creative industries.

Over the last few decades, society has amassed an enormous amount of digital information about the Earth and its inhabitants. However, these archives pale in comparison to the flood of data which is engulfing us. A new wave of technological innovation is allowing us to capture, store, process and display an unprecedented amount of geo-referenced information about our planet and a wide variety of environmental and cultural phenomena. It is today possible to explore, at home, a virtual representation of our planet, flying

* Further author information: (Send correspondence to A.D & M.I.)

A.Doulamis: E-mail: adoulam@cs.ntua.gr, Telephone: +30 28210 37430
M.Ioannides: E-mail: marinos.ioannides@cut.ac.cy, Telephone: +357 25 00 2020

through landscapes draped with aerial or satellite images, contemplating mountains, valleys and other natural sites. Representations of Cultural Heritage (CH) landmarks and entire cities are now available (Himmelstein et al., 2011).

A crucial driving force for the development and economic growth of the creative industries is ICT technologies. Using innovative IT solutions in growing areas of the creative sector - such as advertising, digital media, gaming, including serious gaming, entertainment, interactive design, cultural and educational services - opens up manifold competitive advantages for research, development and business. For this, a lot of work has been done in recent years in the areas of (i) digitisation, especially for tangible cultural (creative) objects, (ii) augmented reality and cultural experiences tools, (iii) information retrieval and use and (iv) virtual heritage, digital libraries and Web archiving.

The rapid development of 3D visualisation services and their interlinking to other on-line services like Google Search, Google Earth / Maps and Bing has forced digital libraries such as Europeana and UNESCOs Memory of the World to expand their content to include 3D museums objects, archaeological sites and monuments. In such a context, digitisation, preservation and online availability of digitised cultural content has always been a top-level priority research.

Today, Photogrammetry's research outcomes in combination with an enormous increase in the availability of computational resources at an affordable price have produced mature applications in the fields of terrain modelling, reconstruction of archaeological monuments or entire cities, the building of virtual worlds in computer games, etc. Moreover, passive scanning methods based on aerial or satellite photography have also been used for topographic modelling of natural open areas with great success. For example, an automatic large-scale stereo reconstruction of urban areas from remote sensing imagery (overlapping aerial images) is described in (Kuschik, 2013). Also, active scanning techniques, such as laser and acoustic methods, take advantage of newer technology advances, accurate geo-referencing and powerful computer machinery. They can produce very dense and accurate 3D point clouds, (Zlatanova, 2008) especially useful for the automatic extraction of height/depth information.

In contrast, Structure from Motion (SfM) algorithms (Hartley and Zisserman, 2004) use data from unstructured sources, e.g. photographs taken by off-the-shelf cameras at different resolutions, time periods, by different people, from various angles and positions in order to recover the structure of a scene as a collection of 3D points as well as the camera position. The problem becomes more complicated and challenging when the source images or videos are harvested from the web as limited or no assumptions can be made about camera characteristics, illumination, resolution, geo-location, etc. However, this extra complication is well worth it given the wealth of information available online and for free over the web. Systems for urban reconstruction from video sources have been proposed (Cornelis et al., 2006, Pollefeys et al., 2008). The latter achieves 3D reconstruction in real time. However, the main advantage of using video as opposed to unordered images is the camera location relationship between successive video frames. Much more complex is the reconstruction from unordered and often irrelevant images. Such a system requires computationally intensive pre-processing and filtering of millions of images using algorithms for landmark identification, for example (Lowe, 2004, Li et al., 2008). These algorithms mainly operate on a pair-wise basis which although time-consuming, it can be massively parallelised, for example (Agarwal et al., 2011). However, the accuracy and the completeness of the point clouds produced by this approach depend highly on the selected images as well as the stereo models used in the multi-view stereo algorithms (Wenzel et al., 2013).

One of the first systems to apply SfM algorithms to internet photo collections was (Snavely et al., 2006). This system was limited to processing only a few thousands of images. A more successful system (Agarwal et al., 2011) can perform dense modelling from Internet photo collections consisting of millions of images. The speedups obtained are due to some refinements e.g. filtering by early 2D appearance-based constraints, but mainly because of utilising graphics processors and parallelising the computation on multi-core computer architectures.

(Fritsch et al., 2012) describes a method for accurate surface reconstruction where 2 billion 3D points were acquired efficiently with sub-mm accuracy using a compact hand-held camera rig consisting of a number of industrial cameras mounted on a frame in order to acquire images from multiple views at once. The images are processed with an automated software pipeline. This method is particularly suited for close-range CH applications; each shot can produce up to 3.5 million 3D points with a maximum precision of 0.2 mm. They report a case where

2 billion 3D points were acquired efficiently with sub-mm accuracy.

In the last two decades, the use of approaches based on 3D geometry has seen rapid progress in many different areas from digital factories of the future to car, flight and surgical training simulators to 3D maps, 3D TV/Cinema and games but also in CH applications. A recent assessment of 3D reconstruction techniques specifically applied to the field of CH is presented in (Manferdini and Galassi, 2013) as a case study focused in Piazza degli Ariani in Ravenna, Italy. (Fassi et al., 2013) compares laser scanning and photogrammetry techniques when applied to a Roman thermal complex in Naples (500m²), the Sant'Erasmus Facade (455 m²) and the Mantegna chapel interior (200m²). It reports that Photogrammetry offers a fully automatic, low cost, fast and accurate method but it requires huge computational resources (and time) when the required accuracy is high. In such cases laser scanning remains the best alternative providing full resolution scans in real time.

In general, linking 3D data (from museums' objects, artefacts, archaeological sites or monuments) with other related 3D objects or with multimedia data, poses different challenges but also provides new, exciting and innovative possibilities compared to more established data forms like texts, images or sound. Subsequently, the digitisation and 3D content creation process in the field of CH is confronted also with another challenge. Each digitised object (geometry) needs logical (semantic) information. Without this logical information, scanned 3D objects like monuments are nothing more than a structure of geometrical primitives (points or triangles). The importance of 4D urban environment modelling can be further emphasised by examining its two main components: (i) 3D content and (ii) time-varying content. (4D World: X+Y+Z+Time=4D); 3D content is of primary importance for our lives as we are living in 3D world and we perceive most of the events occurring in this world by depth information. The importance of 3D content can be highlighted by examining the work of (Fodor, 1983) about the modularity of mind. In this research, we address the aforementioned difficulties by proposing a complete approach able to search, process, model, 3D reconstruct, define metadata and build new production services and re-use of the CH content. In this respect, the rapidly evolving fields of computer graphics, vision and learning, 3D modelling and virtual reality (VR), seems to hold a strong potential for proving effective solutions towards accurate and veridical spatial-temporal urban environment modelling. However, going beyond the state of the art in scene understanding and 3D modelling will enable fundamental new methodologies to view and experience historical and/or temporally varying imagery. This way, we construct building time-varying 3D models that contain how the appearance, historical events around a place and structure have been evolved.

2. PREVIOUS WORK

For 3D scene reconstruction and flow estimation our approach is based on computer vision and Photogrammetry techniques that require optical flow, as well as, disparity estimation. Both can be formulated as a challenge of finding corresponding points in different images. However, the correlation /matching engine is a very challenging task, because a scene's objects, generally, have different shape and appearance when seen from different points of view and at different times.

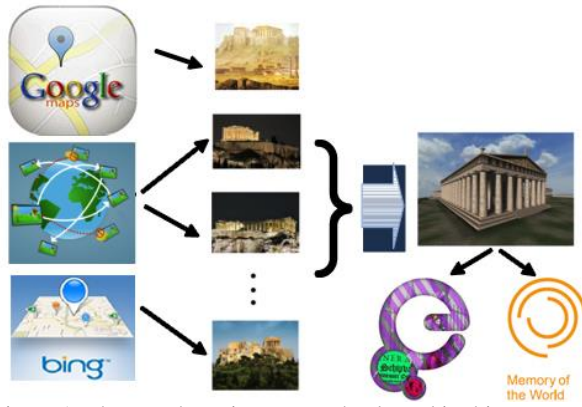


Figure 1: The search engine approach adopted in this paper

The techniques used to solve correlations are similar and can be categorised as energy-based, feature-based and area-based methods. Energy-based methods (Alvarez, 2002, Papenberg et al., 2006) minimise a cost function plus a regularisation term, as in the framework of (Horn and Schunck, 1981), to solve for the 2D displacements. Feature-based methods match features in different images. Selected features should have good discriminative properties and there should be high probability that the same point is selected in different images (Tuytelaars and Mikolajczyk, 2008). This kind of methods are using the concept of coarse-to-fine image warping, however, this down-sampling removes information that may be vital for establishing the correct matches. (Shrivastava et al., 2011) proposes an algorithm, based on the Histogram of Oriented Gradients (HoG) descriptor and the notion of “data-driven uniqueness” by which each image decides what is the best way to weigh its constituent parts, is capable to handle cross-domain image matching. The SIFT flow algorithm presented in (Liu et al., 2008) matches densely-sampled SIFT features between different images, while preserving spatial discontinuities. In (Pons et al., 2007) an approach that overcomes common assumption of feature-based methods limitations is presented. This approach computes a global image-based matching score between different images, making the matching process capable to handle projective distortions and partial occlusions.

3. VISUAL SEARCH ENGINE

In Figure 1, we present a schematic diagram of our search engine (Eagle4D) used for visual content retrieval appropriate for CH objects. The engine is applied on the web and thus it searches objects in the “wild” visual content. Then, the results are improved by exploiting metadata from Europeana and/or Memory of the World.

In this section, we propose a recursive methodology for estimating corresponding points from a set of unstructured and un-calibrated images. In particular, the method exploits local descriptors for finding the most prominent points in an image, which then are used as vertex of a graph. In this way, we construct a graph, the vertices of which coincides with the feature points while the edges the distance between two selected features. In our approach, geometrically invariant features have been selected and especially the SIFT transform. Then, graph partitioning methods are incorporated to estimate those sub-graphs that best match with the respective sub-graphs obtained by images of different view point. The sub-graphs are correlated with each other using graph matching methods. To reduce the computational cost, we introduce recursive approaches for the implementation of the graph partitioning. Particularly, we adopt

the incremental spectral clustering methods, where only a small subset of the constructed graph are updated and a recursive method for estimating the generalized eigenvalues, used for the implementation of the spectral clustering method, have been adopted. Using the matching algorithm, we are able to construct 3D models in a computational efficient approach, resulting in a fast but effective content based retrieval methodology.

Despite the efficiency of the aforementioned methodology, possible errors (in the 3D retrieval process) generate erroneous virtual 3D reconstructions. To address this difficulty, we enhance the results of the 3D matching using metadata information. In particular, we exploit textual information from the set of multi view but unstructured and un-calibrated images in order to allow a first pre-processing. Additionally, using the metadata information, we are able to model the extracted visual descriptors (e.g., the representation of the SIFT features as a graph) with a set of metadata, providing a framework for better categorizing and filtering of the images and thus improving the 3D reconstruction and reducing possible errors.

4. THE VISUAL AUTHORING INTERFACE

The main purpose of this research is to provide the technological framework for enriching the content with additional overlays and virtual objects using smart and cost effective augmented reality methods. The research exploits the outcome of the content based retrieval module (see section 3). This way, the content creators are able to retrieve content of interest in a cost effective manner (through for example the usage of a Sketch or drawing). In the following, the authoring environment tool has the algorithms to transform and scale the retrieved 3D or 4D content in a way to be synchronised with the current scene.

This paper researches on efficient techniques so that the changes in the viewer’s position can be properly reflected in the rendered graphics. Tracking methods are usually based on a two-stage process; a learning or feature extraction stage, and a tracking stage. During the learning stage, some key features relatively invariant to the environment are extracted such that it is mostly a bottom-up process. It can enable re-initialisation of the tracking system to recover when it fails by using automatic recognition. Approaches based on particle filters and Kalman models (Djurić et al. 2003) fail to provide accurate results and on the other to conclude to a computational and cost effective solution. For this reason, in the context of this paper, we incorporate semi-supervised learning algorithms in the area of tracking; Particularly, we allow a minimum user interaction (obtained by the intuitive interactive interface) and then, we implement a classification based tracking strategy to achieve an efficient object tracking and synchronisation.

Personalised Viewing: Since we support a collaborative, interactive interface where multi-users of different properties (e.g., different professions) are incorporated in the authoring process, we need to develop tools for personalised viewing; this means that a part of a scene is presented to the end users according to their particularly interest. In this way, we increase the level of interaction of the content creators, since the system automatically focuses on parts of the scene that are interested by the user.

Perceptual and Real-time Rendering: Rendering process is very expensive and not cost effective. Perpetual rendering transforms rendering into a cost-effective process, since

rendering is implemented in a selective way according to the current content creators interest. Perceptual is required for achieving a cost effective perceptually-based selective rendering algorithms are adopted in order to optimise the 3D world and achieve fast real-time interaction. In order to economise on rendering computation, selective rendering guides high level of detail to specific regions of a synthetic scene and lower quality to the remaining scene, without compromising information transmitted. Such decisions go beyond current task-based approaches and will be guided by context based information.

5. SEMANTIC ENRICHMENT OF CULTURAL HERITAGE ARTEFACTS

A 3D digitised CH artefact is typically composed of geometry and texture data, yet without being semantically enriched with provenance data and its relation to further artefacts as well as with multimedia documentation, it is of very limited practical use.

Therefore the digital representation of an artefact should consist of a geometric structure, accompanied by an annotation to associate semantics and context with its geometry or parts of its 3D shape. The idea of representing semantically structured information and knowledge as well as creating links between the data has increasingly gained popularity, driven by the Semantic Web technologies. Within the current research on annotations, most examples of structured information include semantic models for describing the intrinsic structure of the 3D shape (De Floriani et al., 2008, Papaleo and Floriani, 2009, Attene et al., 2007, Robbiano et al., 2007). (Doerr and Theodoridou, 2011) proposed a model for describing the provenance (life-cycle) of digital 3D shapes in the CH domain (Rodriguez-Echavarría et al., 2009, Havemann et al., 2009, Serna et al., 2011) The Integrated Viewer Browser (IVB) is an interactive semantic enrichment tool for 3D CH collections developed within the 3DCOFORM project. It is fully based on the CIDOC-CRM (ISO 21127:2006) schema supporting its sophisticated annotation model and represents a first approach to a 3D centred and distributed annotation tool taking into account aspects about layout designs, interaction metaphors, and workflows in real professional environments (see figure 2).

Artefacts handled by the Integrated Viewer Browser (IVB) can already be accessed through Europeana, the European Digital Library Portal (see figure 3).

The IVB enables fusion and annotation of a variety of multimedia context information from different sources belonging to a digitised 3D artefact, consolidation of all pieces of information and export of the datasets to Europeana using its ESE (Europeana Semantic Elements) metadata format, a Dublin Core-based set of fields with 12 additional specific Europeana elements to display records appropriately. This semantic enrichment tool will be the focus of future research for further development in order to include the 4th dimension (time) as well as further details related to the origin of the data and the historical value of the object (depending on the application for the re-use of the final 3D results).



Figure 2: General Graphical User Interface of our semantic enrichment tool divided by 5 different sections - relation annotation between Maennerkopf bust and the Saalburg fort, represented by an image of its main entrance

6. CONCLUSION

In this paper we describe a new approach which enables us to reconstruct CH content in 4D, i.e., the 3D spatial reconstruction plus the time. The paper adopts new schemes for 4D visual content reconstruction under a scalable way, visual search at the wild and tools for visualisation and augmented reality. We also describe a visual authoring interface and algorithms for interoperable description of the content and the exploitation of proper metadata adhering to the EUROPEANA framework.

6.1 Acknowledgements

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7-PEOPLE 2007-2013 under REA grant agreement IAPP2012 no 324523.

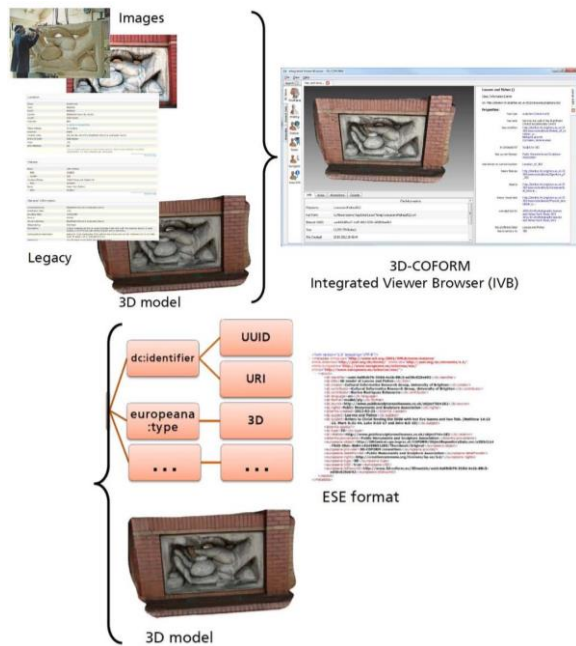


Figure 3: Annotating 3D artefacts with IVB and export to Europeana

7. REFERENCES

References from Journals:

Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M. and Szeliski, R., 2011. Building Rome in a Day. *Communications of the ACM* 54(10), pp. 105–112.

Papenbergh, N., Bruhn, A., Brox, T., Ditas, S. and Weickert, J., 2006. Highly accurate optic flow computation with theoretically justified warping. *INTERNATIONAL JOURNAL OF COMPUTER VISION* 67, pp. 141–158.

References from Proceedings and Books:

Alvarez, L., 2002. Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach. *Journal of Visual Communication and Image Representation* 13(1-2), pp. 3–21.

Attene, M., Robbiano, F., Patan, G., Mortara, M., Spagnuolo, M. and Falcidieno, B., 2007. Semantic annotation of digital 3D ob-jects. In: R. Albertoni and M. Mortara (eds), *Poster and Demo Proceedings of the 2nd International Conference on Semantic and Digital Media Technologies*, Genoa, Italy, December 5-7, 2007, CEUR Workshop Proceedings, Vol. 300, CEUR-WS.org.

Balet, O., 2007. Inscape: an authoring platform for interactive storytelling. In: *Proceedings of the 4th international conference on Virtual storytelling: using virtual reality technologies for storytelling*, ICVS'07, Springer-Verlag, pp. 176–177.

Canciani, M., Falcolini, C., Saccone, M. and Spadafora, G., 2013. From point clouds to architectural models: Algorithms for shape reconstruction. *ISPRS -International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5/W1*, pp. 27–34.

Cornelis, N., Cornelis, K. and Van Gool, L., 2006. Fast compact city modeling for navigation pre-visualization. In: *Proceedings*

of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition -Volume 2, CVPR '06, IEEE Computer Society, pp. 1339–1344.

De Floriani, L., Papaleo, L. and Carissimi, N., 2008. A Java3D framework for inspecting and segmenting 3D models. In: *Proceedings of the 13th international symposium on 3D web technology*, Web3D '08, ACM, pp. 67–74.

Doerr, M. and Theodoridou, M., 2011. CRMDIG: A Generic Digital Provenance Model for Scientific Observation. In: *Proceedings of the 3rd USENIX Workshop on the Theory and Practice of Provenance*, TAPP'11.

Djurić, P.M., Kotecha, Jayesh H., Jianqui Zhang, Yufei Huang, 2003. Particle Filtering, *Signal Processing Magazine*, 33(5), pp. 19-38.

Fassi, F., Fregonese, L., Ackermann, S. and De Troia, V., 2013. Comparison between laser scanning and automated 3D modelling techniques to reconstruct complex and extensive cultural heritage areas. *ISPRS -International Archives of the Photogrammetry, Re-mote Sensing and Spatial Information Sciences XL-5/W1*, pp. 73–80.

Florida, R., 2002. *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books.

Fodor, J., 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Bradford Bks, A BRADFORD BOOK.

Fritsch, D., Abdel-Wahab, M., Cefalu, A. and Wenzel, K., 2012. Photogrammetric point cloud collection with multi-camera 4th systems. In: *Proceedings of the international conference on Progress in Cultural Heritage Preservation*, EuroMed'12, Springer-Verlag, pp. 11–20.

Furukawa, Y. and Ponce, J., 2010. Accurate, Dense, and Robust Multiview Stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32(8), pp. 1362–1376.

Gall, J., Yao, A., Razavi, N., Van Gool, L. and Lempitsky, V., 2011. Hough forests for object detection, tracking, and action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(11), pp. 2188–2202.

Gong, F. and Kang, C., 2009. 3D mesh skeleton extraction based on feature points. In: *Proceedings of the 2009 International Conference on Computer Engineering and Technology -Volume 01*, ICCET '09, IEEE Computer Society, pp. 326–329.

Hartley, R. I. and Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Second edn, Cambridge University Press.

Havemann, S., Settgast, V., Berndt, R., Eide, O. and Fellner, D. W., 2009. The arrigo showcase reloaded – towards a sustainable link between 3D and semantics. *J. Comput. Cult. Herit.* 2(1), pp. 4:1– 4:13.

Hesmondhalgh, D., 2007. *The Cultural Industries*. SAGE Publications.

Himmelstein, J., Balet, O., Ganovelli, F., Gobbetti, E., Specht, M., Mueller, P., Engels, C., Gool, L., de la Riviere, J.-B. and

- Cavazzini, A., 2011. The v-city project: towards rapid and cost-effective reconstruction, visualization and exploitation of complete, large-scale and interactive urban environments. In: The 12th International Symposium on Virtual Reality, Archaeology and Cultural Heritage.
- Horn, B. K. P. and Schunck, B. G., 1981. Determining optical flow. *ARTIFICIAL INTELLIGENCE* 17, pp. 185–203.
- ISO 21127:2006. Information and documentation -- A reference ontology for the interchange of cultural heritage information.
- Kuschik, G., 2013. Large scale urban reconstruction from remote sensing imagery. *ISPRS -International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5/W1*, pp. 139–146.
- Li, X., Wu, C., Zach, C., Lazebnik, S. and Frahm, J.-M., 2008. Modeling and recognition of landmark image collections using iconic scene graphs. In: D. Forsyth, P. Torr and A. Zisserman (eds), *Computer Vision, ECCV 2008, Lecture Notes in Computer Science*, Vol. 5302, Springer-Verlag, pp. 427–440.
- Liu, C., Yuen, J., Torralba, A., Sivic, J. and Freeman, W. T., 2008. SIFT flow: Dense correspondence across different scenes. In: *Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV '08*, Springer-Verlag, pp. 28–42.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, pp. 91–110.
- Manferdini, A. M. and Galassi, M., 2013. Assessments for 3D reconstructions of cultural heritage using digital technologies. *IS-PRS -International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5/W1*, pp. 167–174.
- Mathias, M., Martinovic, A., Weissenberg, J. and Gool, L. V., 2011. Procedural 3D building reconstruction using shape grammars and detectors. In: *Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT '11*, IEEE Computer Society, pp. 304–311.
- Papaleo, L. and Floriani, L., 2009. Semantic-based segmentation and annotation of 3D models. In: *Proceedings of the 15th International Conference on Image Analysis and Processing, ICIAP '09*, Springer-Verlag, pp. 103–112.
- Pollefeys, M., Nistér, D., Frahm, J. M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S. J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G. and Towles, H., 2008. Detailed real-time urban 3D reconstruction from video. *Int. J. Comput. Vision* 78(2-3), pp. 143–167.
- Pons, J.-P., Keriven, R. and Faugeras, O., 2007. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *Int. J. Comput. Vision* 72(2), pp. 179–193.
- Prasad, M., Fitzgibbon, A., Zisserman, A. and Van Gool, L., 2010. Finding nemo: Deformable object class modelling using curve matching. In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 1720–1727.
- Pritchett, P. and Zisserman, A., 1998. Matching and Reconstruction from Widely Separated Views. In: R. Koch and L. Gool (eds), *3D Structure from Multiple Images of Large-Scale Environments, Lecture Notes in Computer Science 1506*, Springer-Verlag, pp. 78–92.
- Pu, S., Vosselman, G. and Vi, C., 2006. Automatic extraction of building features from terrestrial laser scanning. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 33–39.
- Robbiano, F., Attene, M., Spagnuolo, M. and Falcidieno, B., 2007. Part-based annotation of virtual 3D shapes. In: *Proceedings of the 2007 International Conference on Cyberworlds, CW '07*, IEEE Computer Society, pp. 427–436.
- Rodriguez-Echavarría, K., Morris, D. and Arnold, D., 2009. Web based presentation of semantically tagged 3D content for public sculptures and monuments in the UK. In: *Proceedings of the 14th International Conference on 3D Web Technology, Web3D '09*, ACM, pp. 119–126.
- Roussou, M., Laia, P.-T., Olivier, B., Stavrina, P. and Niki, D., 2012. Personalizing interactive digital storytelling in archaeological museums. In: *Proceedings of Computer Applications and Quantitative Methods in Archaeology*.
- Serna, S. P., Scopigno, R., Doerr, M., Theodoridou, M., Georgis, C., Ponchio, F. and Stork, A., 2011. 3D-centered media linking and semantic enrichment through integrated searching, browsing, viewing and annotating. In: *Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage, VAST'11*, Eurographics Association, pp. 89–96.
- Shrivastava, A., Malisiewicz, T., Gupta, A. and Efros, A. A., 2011. Data-driven visual similarity for cross-domain image matching. *ACM Trans. Graph.* 30(6), pp. 154:1–154:10.
- Snively, N., Seitz, S. M. and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph.* 25(3), pp. 835–846.
- Tuytelaars, T. and Gool, L. V., 2000. Wide baseline stereo matching based on local, affinely invariant regions. In: *In Proc. BMVC*, pp. 412–425.
- Tuytelaars, T. and Mikolajczyk, K., 2008. Local invariant feature detectors: a survey. *Found. Trends. Comput. Graph. Vis.* 3(3), pp. 177–280.
- Vitter, J. S., 2001. External memory algorithms and data structures: dealing with massive data. *ACM Comput. Surv.* 33(2), pp. 209–271.
- Wenzel, K., Rothmel, M., Fritsch, D. and Haala, N., 2013. Image acquisition and model selection for multi-view stereo. *ISPRS -International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5/W1*, pp. 251–258.
- Zlatanova, S., 2008. Working group II -acquisition -position paper: Data collection and 3D reconstruction. In: P. Oosterom, S. Zlatanova, F. Penninga and E. Fendel (eds), *Advances in 3D Geoinformation Systems, Lecture Notes in Geoinformation and Cartography*, Springer-Verlag, pp. 425–428.