DIGITAL TECHNIQUES FOR DOCUMENTING AND PRESERVING CULTURAL HERITAGE

Edited by **ANNA BENTKOWSKA-KAFEL** and **LINDSAY MacDONALD**

ARC HUMANITIES PRESS

COSCH

COLOUR & SPACE IN CULTURAL HERITAGE

cost

EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

DIGITAL TECHNIQUES FOR DOCUMENTING AND PRESERVING CULTURAL HERITAGE

COLLECTION DEVELOPMENT,
CULTURAL HERITAGE,
AND DIGITAL HUMANITIES

This exciting series publishes both monographs and edited thematic collections in the broad areas of cultural heritage, digital humanities, collecting and collections, public history and allied areas of applied humanities. In the spirit of our mission to take a stand for the humanities, this series illustrates humanities research keeping pace with technological innovation, globalization, and democratization. We value a variety of established, new, and diverse voices and topics in humanities research and this series provides a platform for publishing the results of cutting-edge projects within these fields.

The aim is to illustrate the impact of humanities research and in particular reflect the exciting new networks developing between researchers and the cultural sector, including archives, libraries and museums, media and the arts, cultural memory and heritage institutions, festivals and tourism, and public history.

DIGITAL TECHNIQUES FOR DOCUMENTING AND PRESERVING CULTURAL HERITAGE

Edited by
ANNA BENTKOWSKA-KAFEL
and **LINDSAY MacDONALD**

ARC HUMANITIES PRESS

MULTIPLE VIEW STEREOVISION

CHRISTOS STENTOUMIS

COSCH Case Study that has applied this technology: Kantharos

Definition

Multiple view, or multi-view stereo vision is the process of reconstructing the 3D model of an object, or scene, from a set of digital images. This 3D model may be generated in several kinds of representation, for example, as a point cloud, as a photo-textured surface, or even as an orthophotograph.

Description

Stereo vision is necessary in the 3D recording of cultural heritage, as it provides accuracy, completeness, and visual quality, as well as cost effectiveness. Thus it plays an important role in the documentation, restoration, preservation, and promotion of cultural heritage assets. Multi-view vision applications may range from small artefacts to large geographic regions by exploiting any kind of optical camera (e.g., SLR, compact, mobile phone) and supporting system (e.g., trolleys, cars, UAVs, robots). Currently, many commercial software and open source algorithms exist for multiple view stereovision; some of these are more generic, others are more application specific. Although such software implementations have brought the 3D product to the public, for more demanding applications the expertise of a professional is still required.

The term multiple view stereovision refers to the automatic reconstruction of a 3D object, or scene (i.e., estimation of 3D coordinates) from more than two source images. Strictly speaking, it could be viewed as essentially an ill-posed 2D-to-3D problem, meaning that the solution is not trivial. This process may involve an arbitrary number of images with different characteristics (positions, rotations, internal camera parameters) from different cameras (still or video frames). The scenes are assumed to be static, thus moving objects are treated as outliers; a different class of methods has been evolved to treat moving and deformable objects.

Model reconstruction from multiple views consists of two main steps: sparse reconstruction and dense reconstruction, while in some cases it can also involve a texture-mapping step. Typically, the process begins with the extraction of interest points, but line or area features may also be exploited. For extracted image points, appropriate characteristic descriptors are calculated, and the point correspondences among images are established via some similarity measure. The retrieved correspondences are used in the bundle adjustment procedure, to restore simultaneously the image orientations (i.e., position and rotations). The Structure from Motion (SfM) process can also provide the camera calibration in multi-view stereo vision.

According to Seitz et al. (2006, 519), "The goal of multi-view stereo is to reconstruct a complete 3D object model from a collection of images taken from known camera viewpoints." Once the full calibration/orientation of each image has been determined, a process of dense image matching estimates all the pixel correspondences with respect to either a putative base image or a world model. A typical taxonomy of multi-view stereo vision algorithms is based on the following dominant criteria: scene representation, photo-consistency measure, visibility model, shape prior, reconstruction algorithm, and initialization requirements (Seitz et al. 2006).

The geometry of an object, or a whole scene, can generally be represented in 3D space via voxels, level-sets, or polygon meshes, which may directly handle issues of visibility and occlusion, and in image projective space via multiple depth maps. The most common photo-consistency measures are the sum of squared intensity differences, which supposes Lambertian surfaces, and cross-correlation, which supposes linear brightness changes. Other more efficient but also more complex matching functions include mutual information, which can handle radiometric differences more effectively, and non-parametric image transformation, which can produce more robust results.

Visibility models are important in multi-view stereo to define which points in the scene are visible and which are occluded in each image. For this purpose, the geometric or photometric attributes of a current 3D reconstructed model and the source images are exploited, thus requiring a two-step process. Moreover, most methods imply, or explicitly exploit, a shape prior regarding the surface of the model. In its simplest form this shape prior is the "fronto-parallel" assumption widely adopted in single stereo-matching methods. The most important classification criterion is the reconstruction algorithm, as this is the field where most advances now occur. Two general categories are global optimization algorithms, such as graph-cuts, level-sets, and PDEs, and algorithms that estimate independent depth maps then fuse them in a full 3D model. Matching algorithms for independent stereo-pairs can also be distinguished as local and global, while between them a group of semi-global and non-local algorithms has also been developed. Finally, the initialization requirements vary among methods, with the most obvi-

ous being the need, or not, of an initial model surface. A multi-view stereovision process can provide several final products, such as photorealistic 3D models, novel viewpoints, and photo-textured mappings, that is, ortho-projections.

The field of multiple view stereovision is rapidly changing as researchers in this area are very active. For a continuously updated list of state-of-the-art algorithms and their performance on dedicated data sets, see one of the various online evaluation platforms, such as http://vision.middlebury.edu/mview/, provided by the Middlebury College, Vermont, USA; The KITTI Vision Benchmark Suite, www.cvlibs.net/datasets/kitti/ of Karlsrue Institute of Technology; and http://cvlab-www.epfl.ch/data/multiview/denseMVS.html by the Computer Vision Laboratory (CVLAB) of the Ecole Polytéchnique Fédérale de Lausanne (EPFL) and part of Strecha et al. (2008).

Significant Applications

Sagalassos

One of the pioneering applications of multi-view stereovision in cultural heritage was the recording of the Sagalassos Hellenic-Roman city in ancient Pisidia (Pollefeys 2002). In this early example of applying computer vision techniques to automatically retrieve the 3D model of a landscape, the accuracy and resolution was far from the standards achieved through photogrammetric techniques, but the process was fully automated. It was possible to retrieve a complete 3D model from images for recording the excavation, the restoration process, or another purpose. Today models are created automatically from mview reconstruction algorithms and software, but the accuracy is not as good as needed, so a human has to correct the model.

Eetioneia Gate

The image-based 3D reconstruction of the East Tower of the Eetioneia Gate is an example of applying multi-view stereo vision algorithms to the recording of cultural heritage. The archaeological site of Eetioneia Gate presents a significant part of the ancient fortification of the Piraeus port of Athens since the fifth century BC. In 2015, the East Tower was recorded in a complete, high-fidelity 3D model, as a part of extensive restoration and conservation activities over the past decade. The tower was captured by 900 images via a 18 MP camera and 17 mm lens from 5 m distance; these resulted in a sub-mm resolution and accuracy of a model without occlusions. Small targets were used to verify the accuracy of the reconstruction. The orientations of the images were solved via SfM and the surface was reconstructed via custom and commercial multi-view stereo vision algorithms. The

products of the contact-less recording, which were useful to architects, archae-ologists, and curators, were a 3D unwrapped surface, orthoprojections viewed from several viewpoints, and the complete 3D photo-textured model. A fly-though video of the model created by up2metric, Athens (www.up2metric.com), is avail-able at https://youtu.be/_w6RA6Xmra0 (accessed 13 February 2017).

Portus Port

The 3D recording of cultural heritage objects has also been studied and evalu-ated on several artefacts in the Portus Project (2007–11) and the Portus in the Roman Mediterranean Project (2011–14) (www.portusproject.org/). Led by the University of Southampton with support from the Soprintendenza Speciale per i Beni Archeologici di Roma and other academic partners, the project aimed at the better understanding of the archaeological site of a large artificial ancient harbour south of Rome. An interdisciplinary approach was adopted, involving the devel-opment of new techniques in all relevant fields: the data capturing, processing, analysis, and presentation of the results. In this scheme, several complementary recording techniques were evaluated in the context of cultural heritage monu-ments. Different means of recording were used, including UAVs, photographs, terrestrial laser scanners, and thermal images.

References

Hartley, R., Zisserman, A. 2003. *Multiple View Geometry in Computer Vision*. Cam-bridge: Cambridge University Press.

Pollefeys, M., Vergauwen, M., Cornelis, K., Verbiest, F., Schouteden, J., Tops, J., van Gool, L. 2002. "3D Acquisition of Archaeological Heritage from Images." *Proc. 18th International CIPA Symposium*, Potsdam, 381–88. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIV-5/C7. ISPRS.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R. 2006. "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms." *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, New York, 519–28. IEEE. doi:10.1109/CVPR.2006.19.

Strecha, C., von Hansen, W., van Gool, L., Fua, P., Thoennessen, U. 2008. "On Bench-marking Camera Calibration and Multi-View Stereo for High Resolution Imagery." *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, 1–8. IEEE. doi:10.1109/CVPR.2008.4587706.

Szeliski, R. 2011. *Computer Vision Algorithms and Applications*. London: Springer. doi:10.1007/978-1-84882-935-0.