

GEOMETRIC AND RADIOMETRIC MODELING OF 3D SCENES

Marco Marcon, Augusto Sarti, Stefano Tubaro

Dipartimento di Elettronica e Informazione - Politecnico di Milano
Piazza Leonardo Da Vinci, 32, 20133 Milano (ITALY)

ABSTRACT

Modeling of 3D scenes is a hot topic in Computer Vision from more than thirty years, and probably its history is longer than a century considering also photogrammetry. In the recent years the rapid technological improvements that characterized the acquisition devices (photo-cameras, video-cameras, ..), illumination devices (lasers, structured light sources) and computational units allowed the application of 3D shape estimation methods, based on image analysis techniques, in a wide set of applications. Furthermore real-time 3D analysis is becoming a common tool in Virtual and Augmented Reality contexts. Aim of this presentation is a rapid description of recent major advances on geometric and radiometric modeling of 3D scenes based on image analysis.

Index Terms— Machine Vision, Geometric modeling, 3D scene reconstruction, Photogrammetry.

1. INTRODUCTION

In general 3D modeling problem formulation is typically connected to a specific application (e.g. biomedical 3D imaging or industrial 3D volume inspection,...) and starts from multiple images captured by analog or digital cameras. It is important to mention that in this type of problems an accurate geometrical description of the framed scene is just one aspect of the modelization procedure. For example considering seamless fusion of real and virtual contents (like those addressed for TV/Film production or for realistic pre-visualization of products, objects, architectural environments as an overlay of a real scene) an accurate description of lighting condition and surface reflection properties of each object present in the scene should be considered. Aspects concerning this topic will be discussed in section 4. In general when a specific 3D modeling application is addressed the preliminary phases that should also be considered can be described as follows:

- definition of the requirements and specifications of the application under investigation;
- sketch of possible solutions or approaches;
- design of an image acquisition model, involving pose planning, sensor design, illumination conditioning etc.;

- test of the defined image acquisition model.

In the following sections we briefly describe recent results on image analysis for 3D scene modeling.

2. ACQUISITION SET-UP

We concentrate our analysis only on passive imaging systems, i.e. systems where no specific lighting devices (like projectors or lasers) are used in the acquisition process. Moreover, in the recent years, some innovative proposals to re-thinking some components of the traditional image acquisition model (e.g. a pinhole projection model with a pre-defined camera motion) led to some interesting results connected to virtual view generation, this approach is normally indicated as "Image Based Rendering" [1][2][3][4]. According to [5] we will divide possible acquisition approaches following this scheme:

- Visual field: *Circular/Non-circular*
- Focal point(s) associated with each image: *Single/Multiple*
- Acquiring time(s): *Single/Multiple*
- Acquiring pose(s): *Single/Multiple*

Most of the classical scene acquisition approaches can be casted into this classification: e.g. a planar image and a cylindrical image are examples of the non-circular and the circular visual field classes. A pinhole projection model is the example for the class of a planar image associated with a single focal point. For the class of a planar (*non-circular*) image associated with multiple focal points many examples are available from multi camera set-ups (bi-nocular or multi-ocular) to images where each pixel is associated with a different focal point, (i.e. orthographic projections from pushbroom cameras [6]). Also image based rendering approaches like Light Field [2] and Lumigraph [1] can be classified in this class where the poses of pinhole cameras are arranged to a planar grid layout and assume synchronous acquisition for their applications. In these approaches the generation of novel views is obtained by a re-sampling of the acquired image data which are parameterized as a 4D function. Approaches based on multiple poses of a pin-hole camera represent the most common case in Computer Vision

and a wide spectrum of processing methodologies are available for different contexts. In many of these approaches a scene geometric modelization is obtained through a set of depth-maps obtained from different view-points and by their fusion. The final result is a complete seamless representation of the scene through space carving [7] [8]. Many examples are also present in literature where time and pose/focus varying acquisition set-ups are used in photometric stereo methods: in [9] orthographic projections with different illumination fields and camera positions at each shot are used for full 3D reconstructions while in [10] different acquisition focuses allows depth-maps reconstruction (depth from De/focusing), a similar approach based on coded aperture and focus sequence is also proposed in [11]. A typical example where multiple view points find useful application is terrain 3D reconstructions. To this aim set-ups where multiple-line scanner systems (pushbroom cameras) are moved over the analyzed surface [6]. Similar considerations also apply to cylindrical images for central-projection panoramas, where different applications are tackled. It is important to notice that even if very significant results has been obtained in the field of 3D reconstruction from images acquired by a single moving camera, the large part of the developed applications for 3D scene reconstruction (both in the computer-vision and in photogrammetry field) are based on the simultaneous use of multiple cameras. This choice is motivated by the fact that the characteristics/performances of the cameras are rapidly growing without significant increase of the costs (or, in same case, with a cost reduction). In the following we will focus on 3D reconstruction typically based on this multiple camera set-up.

3. VOLUMETRIC SCENE RECONSTRUCTION

Volumetric reconstruction of a scene, even if computationally expensive with respect to classical methods based on external surface representation of the imaged scene, represents the most accurate and reliable approach for a large set of practical applications where adherence to the real world volumetric distribution is crucial.

Volumetric data representations have been gaining importance since their introduction in the early 70's in the context of 3D medical imaging [12]. The exponential growth of computational storage and processing power during the last three decades have enabled these representations to become practical alternatives to surface-based geometrical representations for many applications in computer graphics and computer vision [13]. In particular, volumetric models provide a flexible and powerful representation for 3D objects inferred from (typically) multiple images of a scene. Unfortunately there are some differences in the meaning of the word "volumetric" between the disciplines of computer graphics/scientific visualization and that of computer vision. In both cases, the term volumetric implies a representation that describes not only the

external surface of some region, but also the space that the region encloses. However in computer graphics, the term volumetric further implies a sampled representation. Various sampling patterns such as regular, non-isotropic, curvilinear, and unstructured are accommodated, and many of these allow extensions of classical signal and image processing methods to be applied. However, the term volumetric in the field of computer vision implies no such sampling; for example polyhedral representations are considered volumetric in this context. Papers such as [14] and [15], described by vision researchers as involving volumetric representations [16], are examples of non-sampled representations. In this paper we restrict the usage of the term volumetric to imply a sampled space involving the voxel notion. Many volumetric reconstruction techniques described in literature require calibrated input images, which means that the system knows where any 3D point in the scene is projected in each acquired image. Image calibration is itself a challenging problem with a large literature devoted to develop effective estimation algorithms [17] and to its deep connection to scene description and modeling. Calibration methodologies can be divided in 2 principal categories, the first one tries to define intrinsic (internal) camera parameters and extrinsic (the relative position and orientation of each view) using images where a known target is framed. The other technique category, known as *self-calibration*, attempts to find internal and/or external parameters of the employed cameras directly from the available images of the unknown scene, imposing the constraints related to the underlying projective geometry [18].

3.1. The Visual Hull

The earliest attempts at volumetric model reconstruction from photographs are those that approximate the visual hull of the imaged objects. This approach is also referred as "volume intersection" in the literature. The visual hull of an object or a scene, can be described as the maximal shape that gives the same silhouette as the actual object for all views outside the convex hull of the object [19]. In [20] Szeliski gave a substantial improvement to the visual hull building volumetric models directly from acquired images. Szeliski refines a single octree model considering in a sequential way the available images. This allows significant increases in processing speed. Additionally Szeliski is the first to address many practical issues in this context, such as adaptive background subtraction and morphological operations during the segmentation stage. Moreover considering the case of a simple relative motion between the object to be modeled and the camera, he proposed a way for an automatic estimation of that motion. When a large set of images are available, binary segmentations of the considered object(s) can be extracted only on a subset of the considered views, while on each other the segmentation are inferred by computing the trilinear tensor [21] with respect to other two segmented images. An extension of the classical

volume intersection approach is the so called voxel coloring [22]: the algorithm begins with a reconstruction volume of initially opaque voxels that encompasses the scene to be reconstructed. As the algorithm runs, opaque voxels are tested for color consistency and those that are found to be inconsistent are carved, i.e. made transparent. The algorithm stops when all the remaining opaque voxels are color consistent. Further advances concern temporal evolution of the considered 3D model allowing volumetric segmentation based on time evolution, augmented reality and human action tracking and interpretation [23].

A complete different approach is instead based on the representation of a 3D surface as a level-set of an implicit function. Level set theory was originally developed by Osher and Sethian [24] specifically to model the evolution of propagating interfaces. For 3D problems, these methods start with an initial surface which is represented as the set of points where the function $F(x, y, z, t)$ assumes a constant value. The surface then moves along its surface normal with a speed which depends from the local surface properties and from external forces in order to be congruent with the available image data. When the steady-state is reached the surface should be completely consistent with the real world and it will give a 3D model of the imaged scene object(s). Level set methods were initially developed for modeling flame propagation, but have been applied to an astonishingly diverse array of problems [25]. As far as 3D modeling is concerned, different level-set methodologies were applied to steer surface evolution towards the final 3D Object model, in particular corresponding features extracted from multiple calibrated cameras were used to define 3D anchor points [26]; in other applications the surface evolution was steered by a set of 3D points acquired through a laser scanner or other acquisition devices [27].

4. RE-LIGHTING

One of the principal uses of 3D models acquired from images is within TV/Film production and architectural visualization. In these contexts the integration of virtual, synthetic objects into real scenes becoming more and more essential. The final aim is the production of realistic and seamless overlays of virtual objects over real images. Therefore a detailed description of the illumination field that characterize the real scene is very important, due that the fact that also virtual objects should be illuminated in the same way. Methods to capture the scene light field using high dynamic range images (HDRI) were pioneered by P. Debevec [28] and they are now widely used in real productions. These methods involve building up a panoramic lighting representation, by either mapping the environment onto a sphere or a cube. The inserted virtual object is then lit by this HDR illumination map and then overlaid on the real image. Any virtual object needs to have a surface description that gives reflective properties along with the object shape (or geometry) for the rendering system. This might

be the BRDF (bidirectional reflectance distribution function) in the general case or a more simplified characterization based on color, diffuse and specular reflection parameters, are usually assigned manually in the rendering system. The accurate measurement of the BRDF requires a very defined environment. Approaches usually assume a calibrated environment where camera parameters and object shape are precisely known [29] [30]. Further a point light source with known position is used for the computation of the BRDF. Novel techniques try to relax previous constraints, estimating radiometric surface properties using a multi-camera system to obtain an accurate description also for deformable objects (like actors in the scene). This is of paramount importance for the complete relighting of objects/actors, acquired in a studio, that must be rendered in different contexts [31].

5. CONCLUSIONS

This contribution described current state of the art relative to techniques for geometric and radiometric acquisitions of 3D scenes. The final aim was not a complete survey on this field but an overview of the thematic that, in our opinion, have received larger attention of the research community in the last years.

6. REFERENCES

- [1] J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen, "The lumigraph," in *Proceedings of SIGGRAPH 96*, 1996, pp. 43–54.
- [2] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of SIGGRAPH 96*, 1996, pp. 31–42.
- [3] H.Y.Shum and L.W.He, "Rendering with concentric mosaics," in *Proceedings of SIGGRAPH 99*, 1999, pp. 299–306.
- [4] D.N. Wood, A. Finkelstein, J.F. Hughes, C.E. Thayer, and D.H. Salesin, "Multiperspective panoramas for cel animation," in *Proceedings of SIGGRAPH 97*, 1997, pp. 243–250.
- [5] S.K.Wei, F.Huang, and R.Klette, *Classification and Characterization of Image Acquisition for 3D Scene Visualization and Reconstruction Applications*, vol. Volume 2032, chapter Lecture Notes in Computer Science, pp. 81–92, Springer Berlin / Heidelberg, 2001.
- [6] R.Gupta and R.I. Hartley, "Linear pushbroom cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 963–975, 1997.
- [7] G. Dainese, M. Marcon, A. Sarti, and S. Tubaro, "Complete object modeling using a volumetric approach for mesh fusion," in *Proc. 6th International Workshop*

- on *Image Analysis for Multimedia Interactive Services (WIAMIS-2005)*, 2005, pp. 436–444.
- [8] C.Hernandez, G. Vogiatzis, and R. Cipolla, “Multi-view photometric stereo,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 548–554, 2008.
- [9] R. Klette, K. Schluens, and A. Koschan, *Computer Vision - Three-Dimensional Data from Images*, Springer, Singapore.
- [10] P. Rademacher and G. Bishop, “Multiple-center-of-projection images,” in *Proceeding of SIGGRAPH*, 1998, pp. 199–206.
- [11] S. Hiura and T. Matsuyama, “Multi-focus camera with coded aperture: real-time depth measurement and its applications,” in *Proceeding of CDV-WS*, 1998, pp. 101–118.
- [12] J. F. Greenleaf, T. S. Tu, and E. H. Wood, “Computer generated 3-d oscilloscopic images and associated techniques for display and study of the spatial distribution of pulmonary blood flow,” *IEEE Transactions on Nuclear Science*, vol. 17, no. 3, pp. 353, 1970.
- [13] A. Kaufman, *Volume Visualization*, IEEE Computer Society Press, 1991.
- [14] A. Pentland, “Automatic extraction of deformable part models,” *Int.Archives of Photogrammetry and Remote Sensing*, vol. 4, no. 2, pp. 107–126, 1990.
- [15] D. Terzopoulos and D. Metaxas, “Dynamic 3d models with local and global deformations: Deformable superquadrics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 703–714, 1991.
- [16] P. Fua and Y. G. LeClerc, “Object-centered surface representations: Combining multiple-image stereo and shading,” *International Journal of Computer Vision*, vol. 16, no. 1, pp. 35–56, 1995.
- [17] R.I.Hartley and A.Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521623049, 2000.
- [18] E.E.Hemayed, “A survey of camera self-calibration,” in *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003, pp. 351–357.
- [19] A.Laurentini, “The visual hull concept for silhouette-based image understanding,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, vol. 16.
- [20] R. Szeliski, “Rapid octree construction from image sequences,” *Computer Vision, Graphics and Image Processing:Image Understanding*, vol. 58, no. 1, pp. 23–32, 1993.
- [21] A. Shashua and M. Werman, “On the trilinear tensor of three perspective views and its underlying geometry,” in *Proceedings of the IEEE International Conference on Computer Vision*, 1995.
- [22] H. Saito and T. Kanade, “Shape reconstruction in projective grid space from large number of images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1999, vol. 2, pp. 49–54.
- [23] M. Pierobon, M. Marcon, A. Sarti, and S. Tubaro, “3-d body posture tracking for human action template matching,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006*, 2006.
- [24] S. Osher and J. Sethian, “Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations,” *ournal of Computational Physics*, vol. 79, pp. 12–49, 1988.
- [25] J. Sethian, *Level Set Methods and Fast Marching*, Cambridge University Press, 1999.
- [26] S. Tubaro A. Sarti, “Image-based surface modeling: A multi-resolution approach,” *Signal Processing*, vol. 82, no. 9, pp. 1215–1232, 2002.
- [27] Marco Marcon, Luca Piccarreta, Augusto Sarti, and Stefano Tubaro, “Fast pde approach to surface reconstruction from large cloud of points,” *Computer Vision and Image Understanding*, vol. 112, no. 3, pp. 274–285, Dec. 2008.
- [28] P.E.Debevec, “Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography.,” in *Proceedings of SIGGRAPH 98*, 1998, pp. 189–198.
- [29] S.R. Marschner, S.H.Westin, E.P.F. Lafortune, K.E.Torrance, and D.P.Greenberg, “Image-based brdf measurement including human skin,” in *In Proceedings of 10th Eurographics Workshop on Rendering*, 1999, pp. 139–152.
- [30] H.Lensch, J.Kautz, M.Goesele, W.Heidrich, and H.P.Seidel, “Image-based reconstruction of spatial appearance and geometric detail,” in *ACM Transactions on Graphics 22*, 2003, vol. 22, pp. 234–257.
- [31] O. Grau, “Multi-camera radiometric surface modelling for image-based re-lighting,” *Lecture Notes in Computer Science - Springer-Verlag*, vol. 4174, 2006.