

3D Object Reconstruction from Uncalibrated Images using an Off-the-Shelf Camera

Teresa C. S. Azevedo¹, João Manuel R. S. Tavares² & Mário A. P. Vaz³

INEGI^{1,2,3} – Instituto de Engenharia Mecânica e Gestão Industrial

LOME^{1,2,3} – Laboratório de Óptica e Mecânica Experimental

DEMEGI^{2,3} – Departamento de Engenharia Mecânica e Gestão Industrial

FEUP^{1,2,3} – Faculdade de Engenharia da Universidade do Porto – Portugal

E-mail: teresa.azevedo@fe.up.pt¹, tavares@fe.up.pt², gmavaz@fe.up.pt³

Abstract Three-dimensional (3D) objects reconstruction using just bi-dimensional (2D) images has been a major research topic in Computer Vision. However, it is still a hard problem to address, when automation, speed and precision are required and/or the objects have complex shapes or image properties. In this paper, we compare two Active Computer Vision methods frequently used for the 3D reconstruction of objects from image sequences, acquired with a single off-the-shelf CCD camera: Structure From Motion (SFM) and Generalized Voxel Coloring (GVC). SFM recovers the 3D shape of an object based on the relative motion involved, while VC is a volumetric method that uses photo-consistency measures to build the required 3D model. Both methods considered do not impose any kind of restrictions on the relative motion involved.

Introduction

Three-dimensional (3D) models built by computational systems are an intensive and long-lasting research problem for the Graphic and Computer

Vision research communities. Since Computer Vision is concerned with the development of computational theories and methods for the automatic extraction of useful information from images, it offers the opportunity to build 3D models directly from real-world scenes with high accuracy and visual realism.

The main goal of this work was to compare two Computer Vision image-based methods commonly used for 3D objects reconstruction: *Structure From Motion (SFM)* and *Generalized Voxel Coloring (GVC)*. This paper starts with an introduction to the state-of-art in 3D reconstruction, describing some commonly used reconstruction methods. After, some emphasis on *SFM* and *GVC* reconstruction methods is given. Then, the followed methodologies are described. After, some of the obtained experimental results are presented. Finally, some conclusions and guidelines for future work are given.

3D reconstruction

Since most 3D reconstruction methods require considerable computational and imaging resources, there is always a trade-off between used hardware and software, computational complexity and results accuracy, realism and processing speed. In the last decades, the explosive growth in computers processing power and memory storage and their continuous reducing price, has enabled the common use of 3D reconstruction solutions in a variety of application fields, such as:

- Industry, for instance, in clothing industry (e.g. [1, 2]), on-line measurements and production line control (e.g. [3, 4]);
- Navigation systems, for example, in autonomous vehicle guidance (e.g. [5, 6]) and pose estimation (e.g. [7, 8]);
- Virtual reality, such as to build virtual actors, objects or environments (e.g. [9, 10]) and augmented/mixed reality (e.g. [11, 12]);
- Biomedicine, in anthropometric studies (e.g. [13, 14]), detection of tumors or other deformations (e.g. [15, 16]), manufacturing of prosthetic devices (e.g., [17, 18]) and surgery planning (e.g. [19, 20]), for example;
- Architecture/archaeology, for instance, in 3D architectural site reconstruction (e.g. [21, 22]) or archeological documentation (e.g. [23, 24]);
- Security systems, like in visual surveillance (e.g. [25, 26]) and biometric or morphologic information retrieval (e.g. [27, 28]).

Methods for 3D Reconstruction

The usually available methods for 3D reconstruction of objects are typically classified into *contact* or *non-contact*, **Fig. 1**.

Contact-based methods can achieve high accuracy levels and are suitable for a wide range of applications. However, these methods involve mechanical movement of a probe device from one measurement point to the next. Consequently, the data acquisition can be very time consuming. Moreover, since the probes collect only a sparse data set from the object to be reconstructed, some of its critical areas might stay unmeasured. Also, the act of scanning the object by touching it can modify or even damage it, in particular if the object involved is very soft. *CMMs (Coordinate Measuring Machine)* capable of measuring objects of large dimensions are very large in size and so somewhat cumbersome to be used in usual production environments. Furthermore, frequently they need to be placed in controlled-environment rooms, for their protection against temperature variation and vibrations, [29].

Nowadays, the generation of a 3D model is mainly achieved by using non-contact image-based methods. These are usually divided into two main groups, [30]:

1. *Active*: methods that require some sort of energy projection (such as, lasers or structured light) or use the relative motion between camera(s) and objects, to obtain 3D information on the objects shape;
2. *Passive*: methods that do not require energy projection or relative motion, and work under ambient illumination.

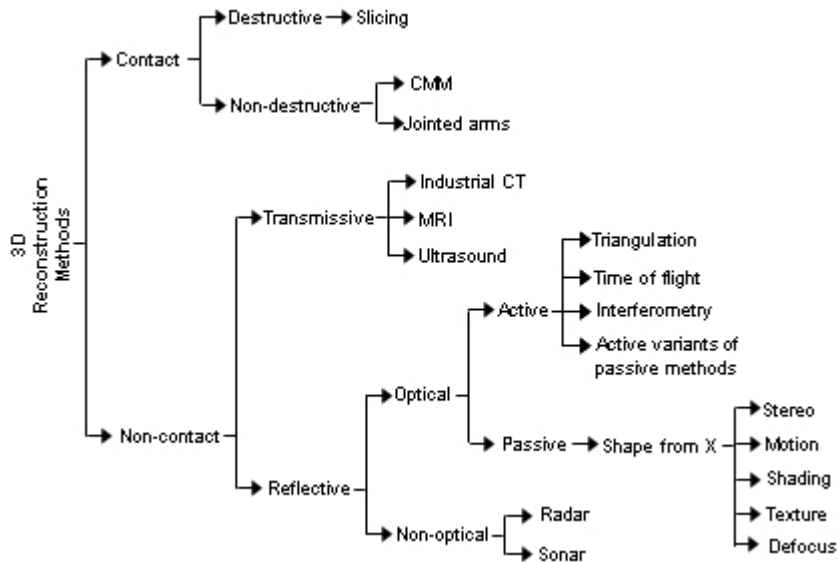


Fig. 1. Common division of usual 3D reconstruction methods

Most common non-contact methods use image data, range sensors, or a combination of both. Image-based methods are widely used, in particular for industrial applications (e.g. [4, 31]), or for precise terrain and city modeling (e.g. [21, 32]).

Range sensors acquire distance measurements from a well known 3D reference coordinate system to the surface points on the object to be reconstructed. They are very common when highly detailed models are required and are already used in industry (e.g. [33, 34]), for documentation of buildings and landscapes (e.g. [35, 36]) or for the recording of objects in archaeology and cultural heritage (e.g. [37, 38]). However, they are costly (at least for now), spatially limited, most of the systems available do not provide color information about the reconstructed object and the quality of the obtained 3D models can be affected by the reflective characteristics of the objects surfaces, [39].

The main difference between image- and range-based methods is that, when using image data, it is necessary to have a mathematical model to derive the objects 3D coordinates, which can be sometimes a complex problem to solve, [40]. Building 3D models using range methods is simpler, because the range data acquired already contains the 3D coordinates necessary for the 3D reconstruction.

The next two subsections will focus on two commonly used image-based reconstruction methods: *Structure From Motion (SFM)*, that belongs

to the standard stereo-based methods, and *Generalized Voxel Coloring (GVC)*, that belongs to the more recent volumetric reconstruction methods.

Structure From Motion

Proposed in [41], *SFM* is a stereo-based method, **Fig. 2**. It uses the relative motion between the camera(s) used and the objects to be reconstructed, to make assumptions about the 3D objects shape. Thus, by knowing the trajectories of objects feature points in the image plane, this method determines the 3D shape and motion that better describes most of the trajectories of the referred points.

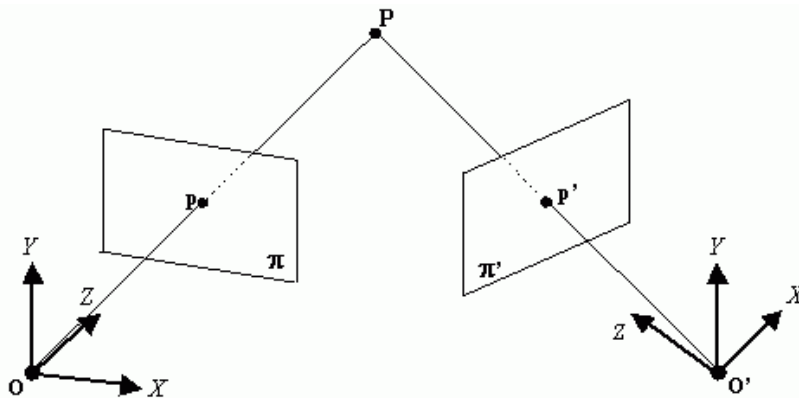


Fig. 2. Stereo vision principle: 3D coordinates of point P are determined through the intersection of the two lines defined by the optical centers O and O' and the matched 2D image points p and p'

This method has received several contributions and diverse approaches: for example, in [42] the use of an extended *Kalman* filter was investigated for estimating the motion and structure from a sequence of monocular images; in [43] an algorithm was developed for shape and motion estimation under orthographic projection using the factorization projection; in [44] a method was proposed that computes the final reconstruction from intermediate reconstructions by analyzing the uncertainties in them, rather than from image data directly; in [45] the problem of solving the *SFM* issue without prior knowledge of point correspondence was addressed; more recently, in [46], an holistic approach was used to compute *SFM* in stages by gradually computing 3D scene information of increasing complexity through processes which operate on increasingly large spatial image areas; among many others. However, *SFM* may suffer from difficulties on finding interest points and/or matching them along the input image sequence, [47]. First, if the object to reconstruct has a smooth surface and low tex-

ture, the extraction of interest features may be difficult or even incorrect since the local appearance is uniform within the neighborhood of each candidate feature. Secondly, matching correspondence cannot be established by just comparing local image measurements, unless the object has a *lambertian* surface; that is, its appearance does not change with the viewpoint. Finally, occlusions in the scene make the correspondence between images difficult or even impossible to obtain.

Generalized Voxel Coloring

As referred earlier, stereo-based methods, like *SFM*, fail to capture objects with complex shapes, smooth surfaces with lack of texture or when occlusion phenomena occur.

For smooth objects, 3D reconstruction using volumetric methods have been quite popular for some time, [48]. These methods are silhouette-based reconstruction methods: intersecting the visual cones generated by the silhouettes and the projection centers of each image, a 3D model can be determined, **Fig. 3**. This 3D model is denominated as *visual hull*, [49], a locally convex over-approximation of the volume occupied by an object.

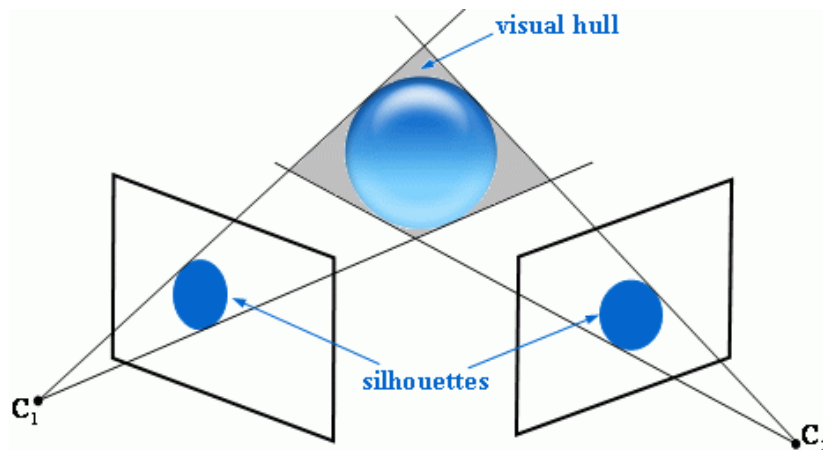


Fig. 3. *Visual hull* obtained from two different viewpoints (C_1 , C_2)

Volumetric methods represent the 3D space model by using *voxels* (regular volumetric structures also known as 3D pixels). The space of interest is divided into discrete voxels which are then classified into two categories: inside and outside. The union of all the inside voxels is an approximation of the visual hull. The accuracy of the reconstruction obtained depends on the number of images used, the positions of each viewpoint

considered, the precision of the camera calibration and the complexity of the objects shape.

Generalized Voxel Coloring (GVC) is a volumetric method that uses photo-consistency criterion, **Fig. 4**, to determine if a certain voxel belongs or not to the object being reconstructed. With this method, the resulting 3D model is the *photo hull*, **Fig. 5**, defined as the largest volume of voxels that are photo-consistent with all viewpoints considered in the reconstruction. Photo-consistency is checked statistically: a voxel is considered consistent if the mean deviation of the pixels color, which results from the voxel image projection, is under a predefined threshold. Thus, *GVC* simultaneously builds and colors the obtained 3D model.

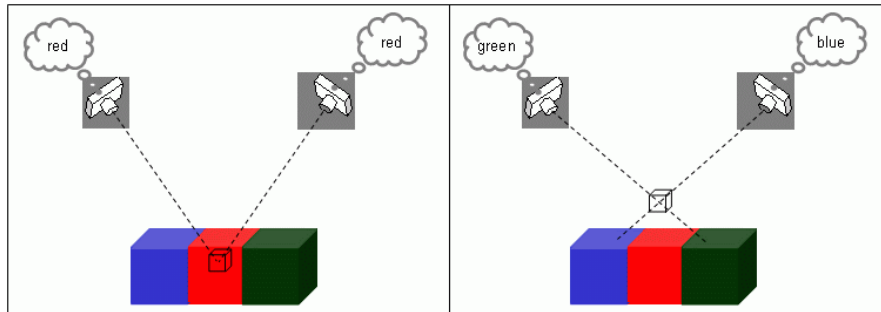


Fig. 4. Color consistency: if the voxel is inside the object surface it will reproject the same color onto all viewpoints where it is visible (left); otherwise, if the voxel is outside the object surface it will most likely reproject distinct colors

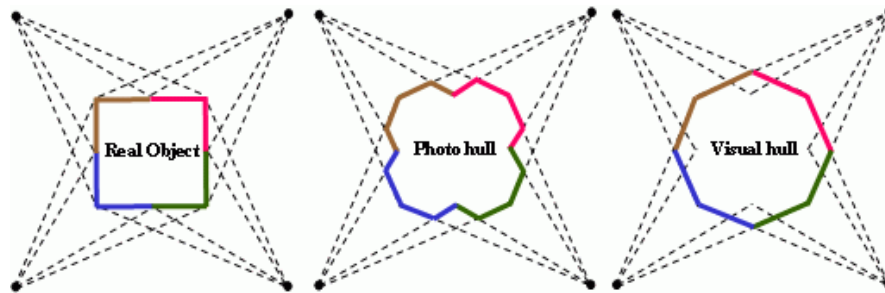


Fig. 5. Relation between *photo* and *visual hull*: the real object is contained inside the *photo hull* that is in turn inside the *visual hull*

Methodologies followed

In this work, *SFM* and *GVC* methods were tested on two objects with different shape properties: a simple parallelepiped and a human hand model.

The parallelepiped has a straightforward topology, with flat orthogonal surfaces, whose vertices are easily detected in each image and simply matched along the acquired image sequence. On the contrary, the hand model has a smooth surface and a more complicated shape.

SFM methodology

To test the *SFM* method, we follow the methodology proposed in [50], and resumed in Fig. 6:

1. the first step is to acquire two uncalibrated images, of the object to be reconstructed, using a single off-the-shelf digital camera;
2. then, image feature points of the considered object are extracted. Feature or interesting points are those who reflect the relevant discrepancies between their intensity values and those of their neighbors. Usually, these points represent vertices, and their correct detection allows posterior matching along the image sequences acquired. Many algorithms for interest points detection are available, but the point features detectors based on the *Harris's* principles, [51], are the most commonly used;
3. after being extracted, feature points must be matched. The matching process is a 2D points association between sequential images that are the projection of the same 3D object point. Automatic detection of matching points between images can be achieved using several cross-correlation processes. They all use small image windows from a first image as templates for matching in the subsequent images, [52]. The most common matching methods include *Normalized Cross-Correlation*, [53, 54], and *Sum-of-Squared-Differences*, [50, 55];
4. then the epipolar geometry is estimated. Epipolar geometry determines a pairwise relative orientation and allows for rejection of previous false matches (or *outliers*). When the interior orientation parameters of both images are the same, it mathematically expresses itself by the *fundamental matrix*, a projective singular correlation between two images, [56]. At least 7 matches are required to compute the fundamental matrix, but to cope with possible outliers, robust methods of estimation are required. In general, the *RANSAC – RANDOM Sampling Consensus* - algorithm, [57], achieves a robust estimation of the epipolar geometry;

5. next step is image rectification. It is the act of projecting two stereo images onto a common plane, such that pairs of conjugate epipolar lines (derived from the fundamental matrix) become collinear and parallel to one of the image axes. Performing this step simplifies the posterior process of dense matching, because the search problem is reduced to 1D;
6. finally, dense matching is performed, where a disparity map is obtained. A disparity map codifies the distance between the object and the camera(s): closer points will have maximal disparity and farther points will get zero disparity. For short, a disparity map gives some perception of discontinuity in terms of depth (2.5D reconstruction).

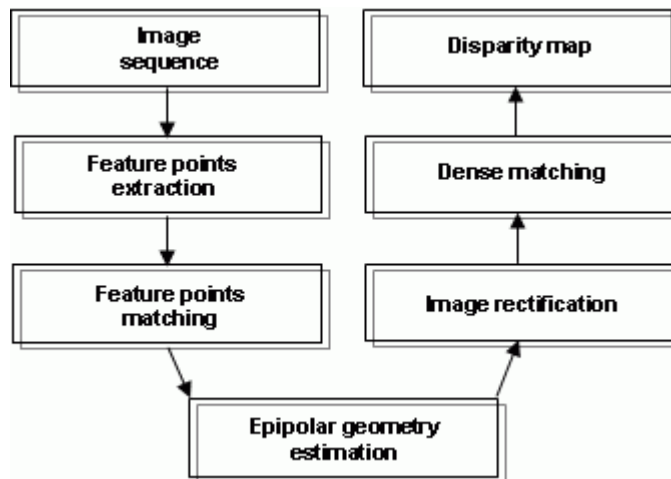


Fig. 6. *SFM* methodology followed for the 3D reconstruction of objects

If the camera were calibrated, the obtained 2.5D reconstruction could be upgraded to 3D, using the triangulation concept, described in Fig. 2.

***GVC* methodology**

To test the *GVC* method we follow the methodology proposed in [58], and represented in Fig. 7.

In this methodology, it is necessary to acquire two image sequences:

- a first one, acquired moving a planar chessboard calibration pattern freely in 3D space;
- for the second sequence, the object to reconstruct is placed on a simple turntable device, with the same chessboard pattern beneath it; keeping the camera untouched, the second sequence of images is acquired, spinning the turntable device until a full rotation is performed.

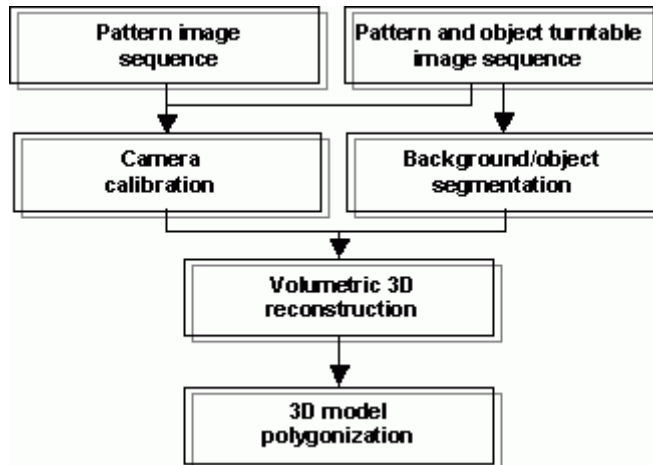


Fig. 7. *GVC* methodology followed for the 3D reconstruction of objects

No restrictions are made on the number of images acquired, nor the rotation angle between two consecutive images of the second image sequence needs to be known.

Then, the used camera is calibrated, in order to find the transformation that maps the 3D world in the associated 2D image space. The calibration procedure is based on *Zhang's* algorithm, [59]. Intrinsic parameters (focal length and principal point) and distortion parameters (radial and tangential) are obtained from the first image sequence; using the second image sequence, the extrinsic parameters (rotation and translation) associated with each viewpoint considered in the reconstruction process are determined.

Then, to obtain the object silhouettes from the input images, image segmentation is performed. This step is required, because, even when the scene background has low color variation, the photo-consistency criterion may not be sufficient for accurate 3D reconstructions, [60]. Also, since the used calibration pattern will rotate along with the object to be reconstructed, it will not be considered as background and, consequently, will be reconstructed as if it was part of the object of interest. Images are here segmented by first removing the red and green channels from the original RGB images and, finally, by image binarization using a user-defined threshold value.

Combining the original image sequence and associated silhouette images, and considering the previously obtained camera calibration parameters, the 3D models are built using the *GVC* volumetric method implemented in [61].

Finally, the volumetric model obtained is polygonized and smoothed using the *Marching Cubes* algorithm ([62]). Basically, this algorithm extracts

a polygonal surface from the volumetrical data. Thus, it proceeds through the voxelized model, and, for each voxel, it determines the polygon(s) needed to represent the patch of the isosurface that passes through the referred voxel.

Experimental results

In this section, some of the obtained experimental results for both followed methodologies and both considered objects will be presented and analyzed.

SFM method

Fig. 8 shows the acquired stereo image pairs of both objects used in this work.

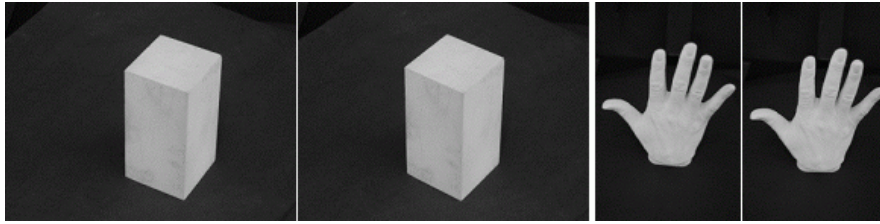


Fig. 8. Stereo image pairs of the objects used to test the *SFM* reconstruction method

For both objects, 200 image features were extracted using the *Harris's* corner detector, [51], imposing a minimum distance between each detected feature. Robust matching of features between the stereo images was made using the *RANSAC* algorithm, [57]. The results obtained can be observed in **Fig. 9**. Since the hand model presents a smooth surface, obviously many wrong matches were detected and, consequently, the determined epipolar geometry will be incorrectly estimated.

After, both stereo pairs were rectified using the algorithm presented in [63]. As observed in **Fig. 10** and **Fig. 11**, the results were much less accurate for the hand model, due to the wrong matches from the previous step. This caused a strong image distortion during the rectification step for this object.

Then, dense matching was performed using *Stan Birchfield's* algorithm, [64]. The results obtained for both objects considered in this work can be observed in **Fig. 12** and **Fig. 13**. Again, from the incorrect results obtained in the previous steps, the dense matching for the hand model was, conse-

quently, of low quality. For the parallelepiped object case, the generated disparity map matches reality better.

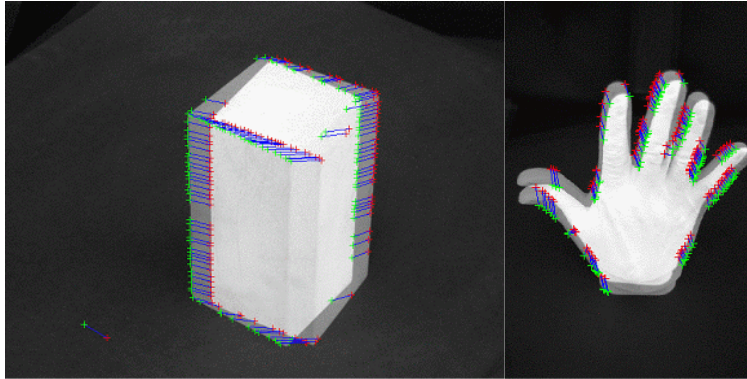


Fig. 9. Results of the (robust) feature points matching for both objects considered: green crosses represent the matched feature points of the first image and the red crosses represent the correspondent matched feature points of the second image

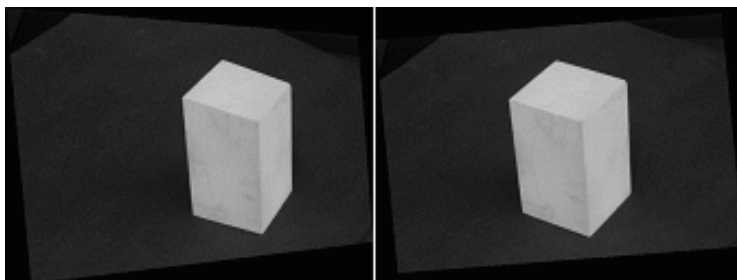


Fig. 10. Rectification results for the stereo images of the parallelepiped object

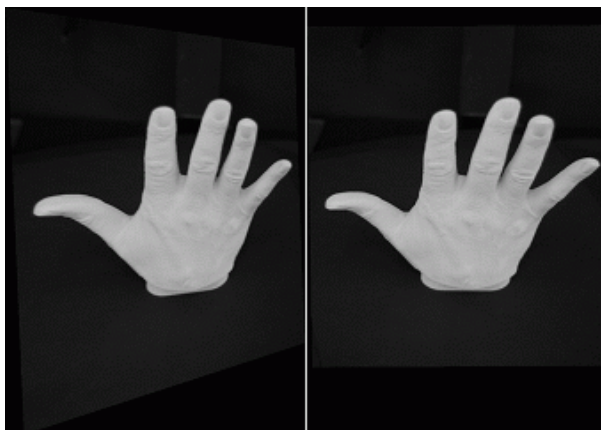


Fig. 11. Rectification results for the stereo images of the hand model object

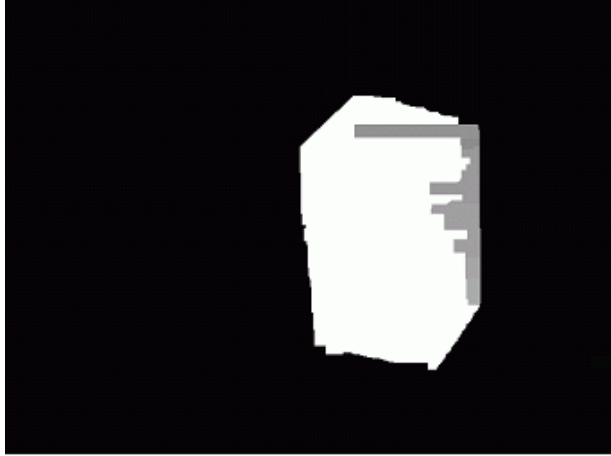


Fig. 12. Disparity map obtained for the parallelepiped object

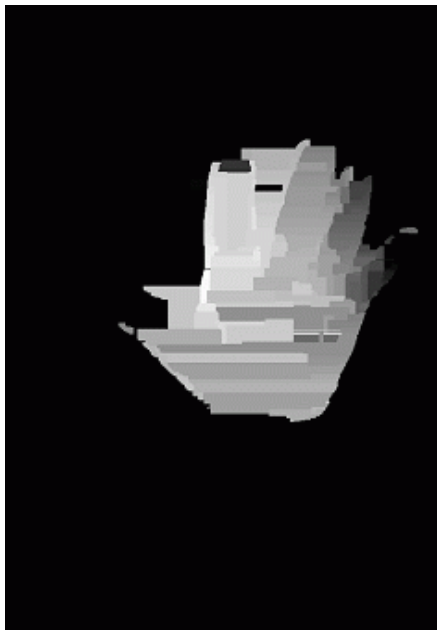


Fig. 13. Disparity map obtained for the hand model object

***GVC* method**

Fig. 14 shows some examples of the second image sequence acquired for the 3D reconstruction of both objects using the *GVC* method.

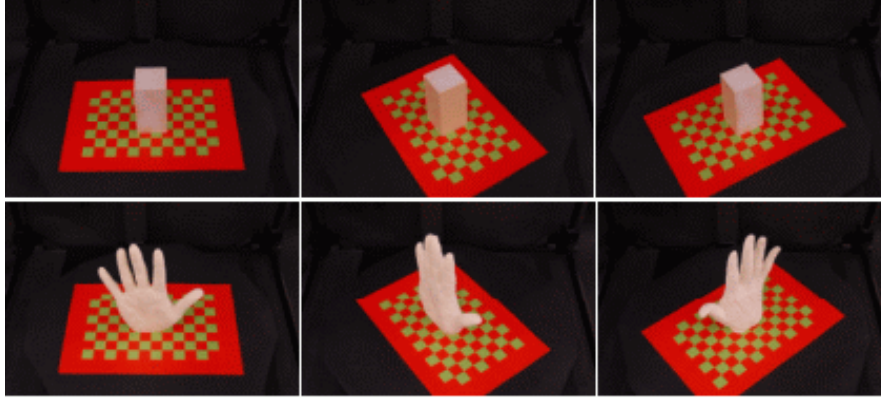


Fig. 14. Three images used for the 3D reconstruction of the parallelepiped (top) and the hand model (bottom)

For both objects considered, the results of the extrinsic calibration procedure are represented in **Fig. 15**. The 3D graphics shown represent the viewpoints considered in the second image acquisition process, considering the world coordinate system fixed on the lower-left corner of the chessboard pattern and the camera rotating around the object.

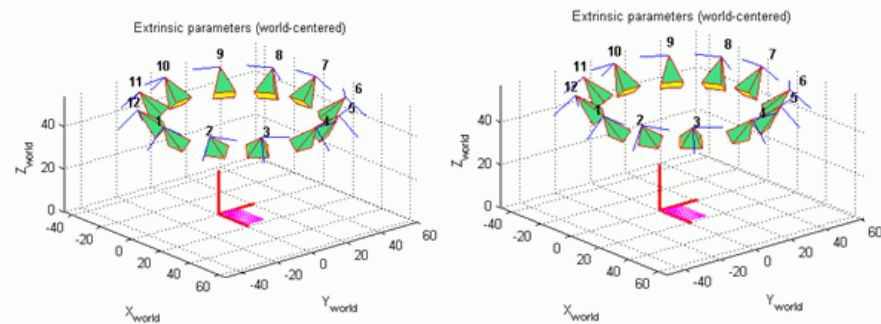


Fig. 15. 3D graphical representation of the extrinsic parameters obtained from the camera calibration process for the parallelepiped object case, on the left, and for the hand model case, on the right

Another way to verify the accuracy of the calibration results obtained is to reproject the 3D points from the chessboard pattern in all images of the second sequence considered. The standard deviations of the reprojection errors (in pixels) for the hand and torso models cases are indicated in **Table 1**. The results obtained from the camera calibration were very accurate for both cases.

The efficacy of our segmentation method enabled us to obtain good silhouette images for both considered objects, **Fig. 16**.

Table 1. Error of the reprojection of the pattern points into all images of the second image sequence

Object	Reprojection error (in pixels)			
	Average		standard deviation	
	x	y	x	y
Parallelepiped	-1.24e-04	-2.67e-05	0.545	0.594
Hand model	-7.31e-05	-2.61e-05	0.673	0.840

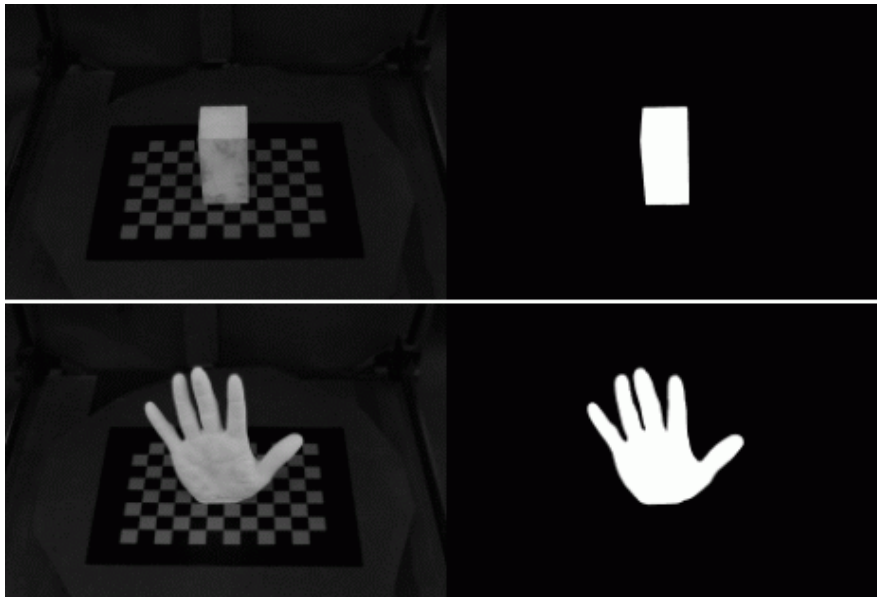


Fig. 16. One example of image segmentation for the parallelepiped (top) and the hand model (bottom): on the left, the original image; on the right, the binary image obtained

Fig. 17 and **Fig. 18** show the results of the 3D reconstruction obtained for both objects using the *GVC* method. Both reconstructed models are very similar to the real 3D object, even in the case of the hand model. Comparing these results with the previous obtained by the *SFM* methodology, *GVC* has no problem to reconstruct objects with smooth and complex shapes. On the other hand, the accuracy of the 3D models built by this last methodology is highly dependent on the calibration and segmentation steps. Thus, *GVC* puts some restrictions, such as a background with low color variation and suitable calibration apparatus, making it less appropriated for unconstrained real-world reconstructions.

GVC methodology was also tested on another object to verify its accuracy: a torso model. Comparing with the previous used objects, the torso

has considerably higher dimensions. Thus, a different calibration pattern was required. Some of the results obtained after the reconstruction process can be seen on **Fig. 19**, where it can be noticed that the torso reflects the calibration pattern on its surface. As consequence, the inferior zone of the reconstructed 3D model is not very accurate, both in terms of shape and color.

From the voxelized 3D model obtained, some geometrical measures can be determined, such as height, length and width. **Fig. 20** compares these values with the real ones, obtained using an usual ruler, for all reconstructed objects. This comparison confirms the approximated reconstruction results of the considered objects, using the *GVC* methodology.

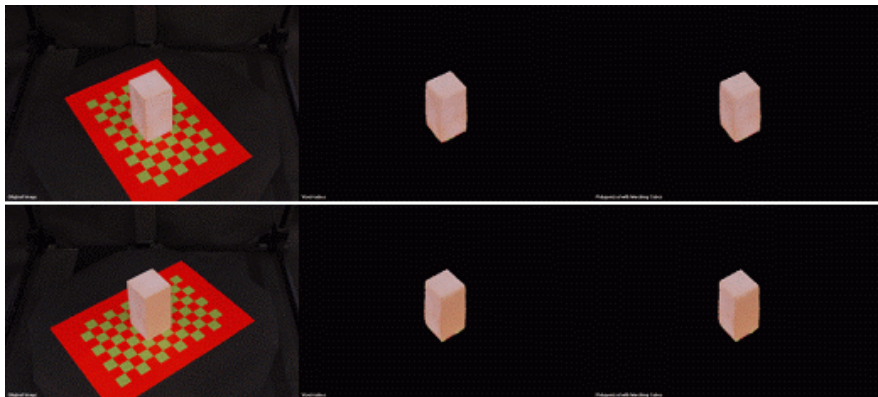


Fig. 17. Two different viewpoints (by row) of the 3D model obtained for the parallelepiped case: on the left, original image; in the centre, voxelized 3D model; on the right, polygonized and smoothed 3D model

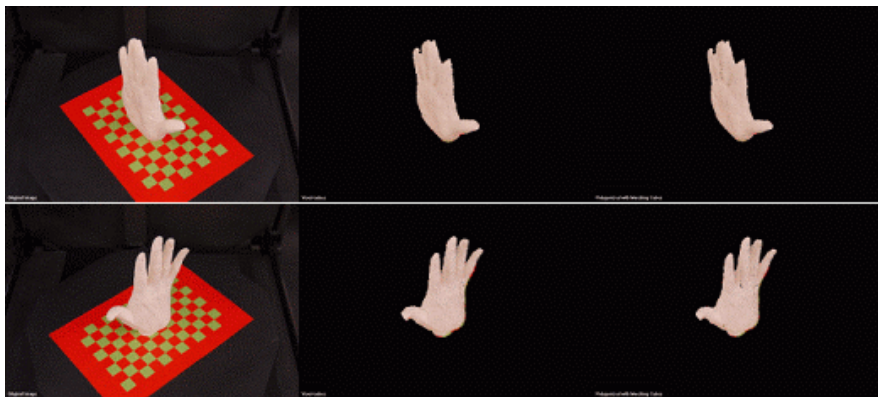


Fig. 18. Two different viewpoints (by row) of the 3D model obtained for the hand model case: on the left, original image; in the centre, voxelized 3D model; on the right, polygonized and smoothed 3D model

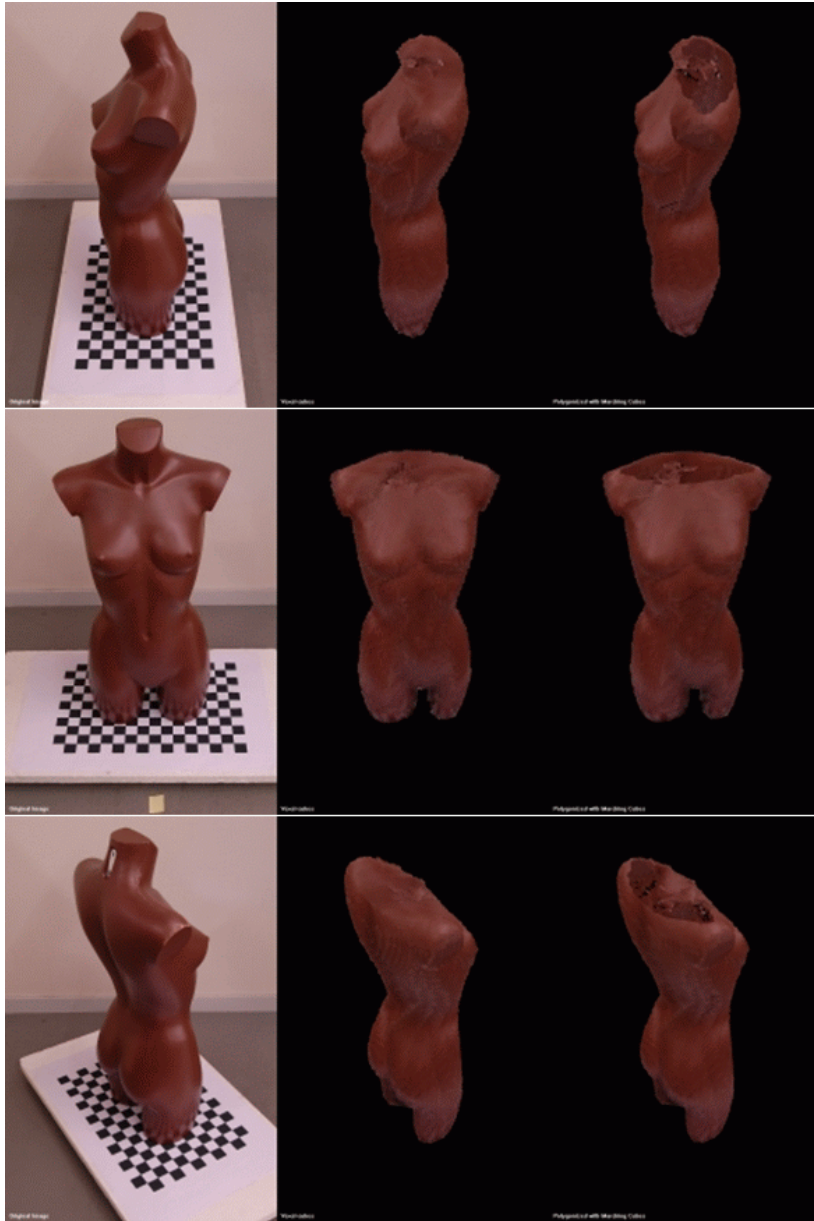


Fig. 19. Three different viewpoints (by row) of the 3D model obtained for the torso model case: on the left, original image; in the centre, voxelized 3D model; on the right, polygonized and smoothed 3D model

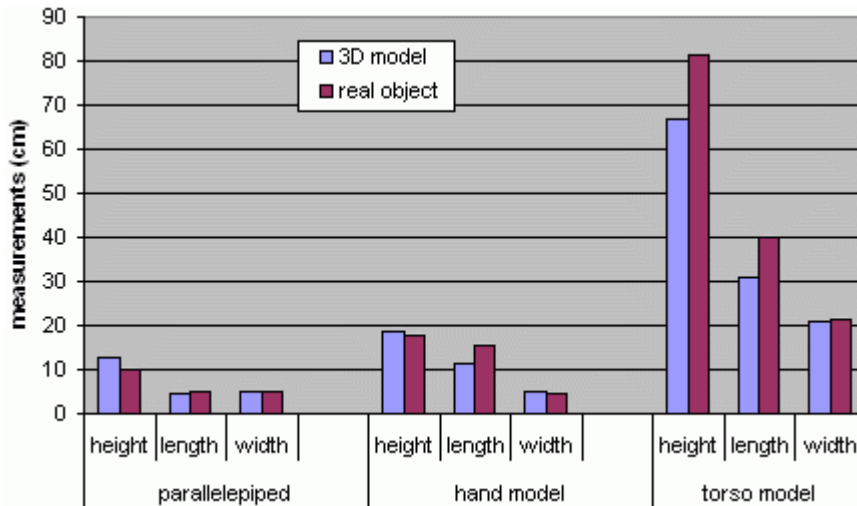


Fig. 20. Comparison of the obtained measurements from the reconstructed 3D models with the real objects measures

Conclusions

The main goal of this paper was to compare experimentally two commonly used image-based methods for 3D object reconstruction: *Structure From Motion (SFM)* and *Generalized Voxel Coloring (GVC)*.

To test and compare both methods, two objects with different shape properties were used: a parallelepiped and a hand model.

Our adopted *SFM* methodology produced fine results when the objects present strong feature points, and so, are easy to detect and match along the input images. However, we can conclude that even small errors in the matching or in the epipolar geometry estimation can seriously compromise the success of the remaining steps.

The models built using the *GVC* methodology were quite similar to the real objects, be it in terms of shape or in color. Nevertheless, the reconstruction accuracy was highly dependent on the quality of the results from camera calibration and image segmentation steps. These can be two major drawbacks in real-world scenes, because they can limit the application of the *GVC* method. Moreover, the reflectance of their surfaces is an aspect that must be considered for more accurate 3D reconstructions. In resume, we can conclude that in controlled environments the *GVC* methodology is capable to obtain adequate 3D static reconstructions of objects from im-

ages. In addition, its major contribution may be the fact that it is fully automatic and suitable for many real applications.

Thus, when comparing the two methods, we can conclude that, on one hand, *GVC* performs better in 3D reconstruction of objects with complex shapes and, on the other hand, *SFM* is better for unconstrained real-world objects reconstruction.

Acknowledgements

This work was partially done in the scope of project “Segmentation, Tracking and Motion Analysis of Deformable (2D/3D) Objects using Physical Principles”, with reference POSC/EEA-SRI/55386/2004, financially supported by *FCT – Fundação para a Ciência e a Tecnologia* from Portugal.

The first author would like to thank the support of the PhD grant SFRH/BD/27716/2006, also from *FCT*.

References

1. T. I. Vassilev (2000). Dressing Virtual People, Systemics, Cybernetics and Informatics, Orlando, FL, USA.
2. P. Volino, F. Cordier and N. Magnenat-Thalmann (2005). From early virtual garment simulation to interactive fashion design, *Computer-Aided Design*, 37(6): 593-608.
3. S. M. Bhandarkar, T. D. Faust and M. Tang (1999). CATALOG: a system for detection and rendering of internal log defects using computer tomography, *Machine Vision and Applications*, 11: 171-190.
4. L. Song, X. Qu and S. Ye (2007). Improved SFS 3D measurement based on BP neural network, *Image and Vision Computing*, 25(5): 614-622.
5. D. Cobzas, H. Zhang and M. Jagersand (2003). Image-based localization with depth-enhanced image map, *IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, 2: 1570-1575.
6. K. O. Arras, N. Tomatis, B. T. Jensen, et al. (2001). Multisensor on-the-fly localization: Precision and reliability for applications, *Robotics and Autonomous Systems*, 34(2-3): 131-143.
7. A. J. Davison, A. G. Cid and N. Kita (2004). Real-time 3D SLAM with wide-angle vision, *5th IFAC Symposium on Intelligent Autonomous Vehicles*, Lisbon, Portugal.
8. T. Lemaire, C. Berger, I.-K. Jung, et al. (2007). Vision-Based SLAM: Stereo and Monocular Approaches, *International Journal of Computer Vision*, 74(3): 343-364.

9. J.-C. Nebel, A. Sibiryakov and X. Ju (2003). V-Man Generation for 3-D Real Time Animation, A Symposium on Intelligent Motion and Interaction Within Virtual Environments, London, UK.
10. A. Koutsoudis, F. Arnaoutoglou and C. Chamzas (2007). On 3D reconstruction of the old city of Xanthi. A minimum budget approach to previous term-virtualnext term touring based on photogrammetry, *Journal of Cultural Heritage*, 8(1): 26-31.
11. H. Aans and F. Kahl (2002). Estimation of Deformable Structure and Motion, *Vision and Modelling of Dynamic Scenes Workshop*, Copenhagen, Denmark.
12. B. Vagvolgyi, C. Reiley, G. Hager, et al. (2008). Augmented Reality using Registration of 3D Computed Tomography to Stereoscopic Video of Laparoscopic Renal Surgery, *The Journal of Urology*, 179(4): 241-242.
13. P. Tukuisis, P. Meunier and C. E. Jubenville (2001). Human body surface area: measurement and prediction using three dimensional body scans, *European Journal of Applied Physiology and Occupational Physiology*, 85(3-4): 264-271.
14. L. Benton and J.-C. Nebel (2002). Study of the breathing pattern based on 4D data collected by a dynamic 3D body scanner, *7th Numérisation 3D/Scanning*, Paris, France.
15. C. Beneder, F. G. Fuechsel, T. Krause, et al. (2008). The role of 3D fusion imaging in sentinel lymphadenectomy for vulvar cancer, *Gynecologic Oncology*, 109(1): 76-80.
16. G. D. Giammarco, M. L. Storto, R. Marano, et al. (2006). Superior vena cava syndrome: a 3D CT-scan reconstruction, *European Journal of Cardio-Thoracic Surgery*, 30(2): 384-385.
17. N. H. Mahmood and T. Tjahjadi (2006). 3D Reconstruction From Multiple Views For Orthotic and Prosthetic Design: An Overview, *SCORED - Student Conference on Research and Development*, Shah Alam, Selangor, Malaysia, 27-28.
18. S. I. Buchillard, S. H. Ong, Y. Payan, et al. (2007). 3D statistical models for tooth surface reconstruction, *Computers in Biology and Medicine*, 37(10): 1461-1471.
19. A. Pednekar and I. A. Kakadiaris (2000). Applications of Virtual Reality In Surgery, *Indian conference on Computer Vision, Graphics and Image Processing, Graphics and Applications session*, Bangalore, India.
20. J. P. Helferty, A. J. Sherbondy, A. P. Kiraly, et al. (2007). Computer-based system for the virtual-endoscopic guidance of bronchoscopy, *Computer Vision and Image Understanding*, 108(1-2): 171-187.
21. D. P. Robertson and R. Cipolla (2002). Building Architectural Models from Many Views Using Map Constraints, *7th European Conference on Computer Vision-Part II*, Copenhagen, Denmark, 155-169.
22. S. Lee, D. Feng and B. Gooch (2008). Automatic construction of 3D models from architectural line drawings, *Symposium on Interactive 3D Graphics and Games*, Redwood City, California, CA, USA, 123-130.
23. R. Sablatnig, S. Tosovic and M. Kampel (2002). Combining shape from silhouette and shape from structured light for volume estimation of archaeological

- cal vessels, International Conference on Pattern Recognition, Quebec City, QC, USA, 1: 364-367.
24. M. Kampel and R. Sablatnig (2007). Rule based system for archaeological pottery classification, *Pattern Recognition Letters*, 28(6): 740-747.
 25. A. Calbi, C. S. Regazzoni and L. Marcenaro (2006). Dynamic Scene Reconstruction For Efficient Remote Surveillance, *IEEE International Conference on Video and Signal Based Surveillance*, Sydney, Australia, 99-99.
 26. S. Fleck, F. Busch, P. Biber, et al. (2006). 3D Surveillance A Distributed Network of Smart Cameras for Real-Time Tracking and its Visualization in 3D, *IEEE Computer Vision and Pattern Recognition, Workshop on Embedded Computer Vision*, New York, NY, USA, 118-118.
 27. N. Lynnerup, M. Andersen and H. P. Lauritsen (2003). Facial image identification using Photomodeler[®], *Legal Medicine*, 5(3): 156-160.
 28. J. Lee, E.-D. Lee, H.-O. Tark, et al. (2008). Efficient height measurement method of surveillance camera image, *Forensic Science International*, 177(1): 17-23.
 29. R. Gershon and M. Benady (2001). Noncontact 3-D measurement technology enters a new era, <http://www.qualitydigest.com/sept01/html/3d.html> (retrieved in June 2007).
 30. S. M. Seitz (1999). An Overview of Passive Vision Techniques, *SIGGRAPH 2000 Course on 3D Photography*, Course Notes, New Orleans, Louisiana, LA, USA.
 31. S. M. Youssef and R. M. Salem (2007). Automated barcode recognition for smart identification and inspection automation, *Expert Systems with Applications*, 33(4): 968-977.
 32. Y. Yu and J. Malik (1998). Recovering Photometric Properties of Architectural Scenes from Photographs, *Computer Graphics*, 32: 207-217.
 33. G. Zhang, J. He and X. Li (2005). 3D vision inspection for internal surface based on circle structured light, *Sensors and Actuators A: Physical*, 122(1): 68-75.
 34. C. P. Witana, S. Xiong, J. Zhao, et al. (2006). Foot measurements from three-dimensional scans: A comparison and evaluation of different methods, *International Journal of Industrial Ergonomics*, 36(9): 789-807.
 35. C. Früh and A. Zakhor (2004). An Automated Method for Large-Scale, Ground-Based City Model Acquisition, *International Journal of Computer Vision*, 60(1): 5-24.
 36. J. Chen and B. Chen (2008). Architectural Modeling from Sparsely Scanned Range Data, *International Journal of Computer Vision*, 78(2-3): 223-236.
 37. M. Levoy, K. Pulli, B. Curless, et al. (2000). The Digital Michelangelo Project: 3D scanning of large statues, *Siggraph 2000*, *Computer Graphics*, ACM Press, ACM SIGGRAPH, Addison Wesley, 131-144.
 38. G. Guidi, B. Frischer, M. Russo, et al. (2006). Three-dimensional acquisition of large and detailed cultural heritage objects, *Machine Vision and Applications*, 17(6): 349-360.
 39. F. Remondino and S. El-Hakim (2006). Image-Based 3D Modelling: A Review, *The Photogrammetric Record*, 21(115): 269-291.

40. F. Remondino, A. Guarnieri and A. Vettore (2005). 3D Modeling of close-range objects: photogrammetry or laser scanning?, *SPIE-IS&T Electronic Imaging*, 5665: 216-225.
41. S. Ullman (1979). *The Interpretation of Visual Motion*, Massachusetts MIT Press, Cambridge, Massachusetts, MA, USA.
42. T. Broida and R. Chellappa (1991). Estimating the Kinematics and Structure of a Rigid Object from a Sequence of Monocular Images, *Pattern Analysis and Machine Intelligence*, 13: 497-513.
43. C. Tomasi and T. Kanade (1992). Shape and motion from image streams under orthography: A factorization method, *International Journal of Computer Vision*, 9: 137-154.
44. J. Oliensis (1999). A multi-frame structure-from-motion algorithm under perspective projection, *International Journal of Computer Vision*, 34: 1-30.
45. F. Dellaert, S. Seitz, C. Thorpe, et al. (2000). Structure from Motion without Correspondence, *IEEE Computer Vision and Pattern Recognition*, Hilton Head, SC, USA, 2: 557-564.
46. J. Hui (2006). A holistic approach to structure from motion, PhD on Computer Science, University of Maryland, MD, USA.
47. R. Hartley and A. Zisserman (2004). *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd ed.
48. S. Seitz and C. R. Dyer (1997). Photorealistic Scene Reconstruction by Voxel Coloring, *Computer Vision and Pattern Recognition Conference*, San Juan, Puerto Rico, 1067-1073.
49. A. Laurentini (1994). The visual hull concept for silhouette-based image understanding, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2): 150-162.
50. M. Pollefeys, L. V. Gool, M. Vergauwen, et al. (2004). Visual Modeling with a Hand-Held Camera, *International Journal of Computer Vision*, 59(3): 207-232.
51. C. G. Harris and M. J. Stephens (1988). A combined corner and edge detector, *Forth Alvey Vision Conference*, University of Manchester, England, 15: 147-151.
52. R. C. Gonzalez and P. Wintz (1987). *Digital Image Processing*, 2nd ed., Addison Wesley.
53. P. J. Burt, C. Yen and X. Xu (1982). Local Correlation Measures for Motion Analysis: A Comparative Study, *Pattern Recognition and Image Processing*, Las Vegas, NV, USA, 269-274.
54. J. P. Lewis (1995). Fast Normalized Cross Correlation, *Vision Interface*, 120-123.
55. A. Moore (2006). *Stereo*, Computer Vision Lectures, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA.
56. O. Faugeras, Q.-T. Luong and S. J. Maybank (1992). Camera self-calibration: Theory and experiments, *2nd European Conference on Computer Vision*, Santa Margherita Ligure, Italy, *Lecture Notes in Computer Vision*, Springer-Verlag, 588: 321-334.

57. M. A. Fischler and R. Bolles (1981). RANdom SAmping Consensus: a paradigm for model fitting with application to image analysis and automated cartography, *Communications of the ACM*, New York, NY, USA, 24(6): 381-395.
58. T. C. S. Azevedo, J. M. R. S. Tavares and M. A. P. Vaz (2007). 3D Volumetric Reconstruction and Characterization of Objects from Uncalibrated Images, 7th IASTED International Conference on Visualization, Imaging, and Image Processing, Palma de Mallorca, Spain, 141-146.
59. Z. Zhang (2000). A Flexible New Technique for Camera Calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11): 1330-1334.
60. K. Sande (2004). A Practical Setup for Voxel Coloring using off-the-shelf Components, Bachelor Project, Universiteit van Amsterdam, Netherlands.
61. M. Loper (2002). Archimedes: Shape Reconstruction from Pictures - A Generalized Voxel Coloring Implementation, <http://matt.loper.org/Archimedes/> (retrieved in August 2006).
62. W. E. Lorensen and H. E. Cline (1987). Marching cubes: A high resolution 3D surface construction algorithm, *ACM SIGGRAPH Computer Graphics*, New York, NY, USA, 21(4): 163-169.
63. F. Isgrò and E. Trucco (1999). Projective rectification without epipolar geometry, *IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, CO, USA, 1: 94-99.
64. S. Birchfield (1999). Depth Discontinuities by Pixel-to-Pixel Stereo, *International Journal of Computer Vision*, 35(3): 269-293, <http://vision.stanford.edu/~birch/p2p/> (retrieved in November 2004).