LaserGun: a Tool for Hybrid 3D Reconstruction

Marco Fanfani and Carlo Colombo

Computational Vision Group
Dip. di Ingegneria dell'Informazione
Universitá di Firenze
Via S. Marta 3, 50139 Firenze, Italy
{marco.fanfani,carlo.colombo}@unifi.it
http://cvg.dsi.unifi.it/

Abstract. We present a tool for the acquisition of 3D textured models of objects of desktop size using an hybrid computer vision framework. This framework combines active laser-based triangulation with passive motion estimation. The 3D models are obtained by motion-based alignment (with respect to a fixed world frame) of imaged laser profiles back-projected onto time-varying camera frames. Two distinct techniques for estimating camera displacements are described and evaluated. The first is based on a Simultaneous Localization and Mapping (SLAM) approach, while the second exploits a planar pattern in the scene and recovers motion by homography decomposition. Results obtained with a custom laser-camera stereo setup — implemented with off-the-shelf hardware — show that a trade-off exists between the greater operational flexibility of SLAM and the higher model accuracy of the homography-based approach.

Keywords: 3D Reconstruction, Active Triangulation, Motion Estimation, SLAM

1 Introduction

Visual reconstruction of 3D object shape is typically accomplished through either active or passive methods.

Active methods rely on the observation of a light pattern while it interacts with the scanned object [1, 2]. Accurate models are obtained also for textureless objects, working in structured conditions with sophisticated hardware and relatively simple algorithms. In [3], the authors use a pattern with several light stripes arranged in a regular way; object shape is obtained through the so called *active triangulation* approach of single image points. In [4, 5] two 3D reconstruction approaches based on active triangulation of a hand-held laser device are described. The former uses a laser blade and requires a background with known geometry to simultaneously estimate the laser plane equation and reconstruct small-size objects. The latter — used also to reconstruct room-size environments — uses an ad-hoc pointer array device and requires an initial calibration step. A

different approach, called *active rectification* and based on image warping transformations of entire laser profiles, is described in [6].
Passive methods use only unstructured illumination, and focus instead on low cost hardware and sophisticated software, by which a reasonable accuracy and a high flexibility can be obtained. Typical passive approaches encompass multi-view reconstruction from either image collections [7, 8] or image sequences [9], real time stereo [10] and shape from shading [11].
In [12], an active/passive method is presented where cast shadows produced with a wand are used instead of projected light. Another hybrid approach extending standard shape from shading is photometric stereo [13], where a collection of photos of the object is taken from a single viewpoint by varying the light source.

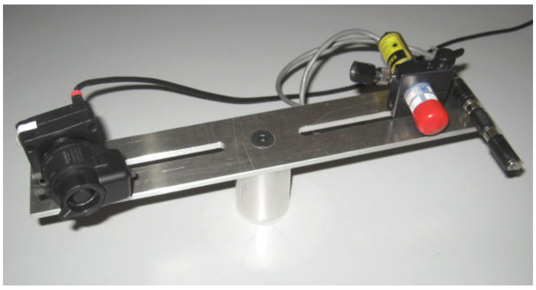


Fig. 1. The camera-laser group for free-hand 3D acquisition.

In this paper, a hybrid solution to the 3D reconstruction problem is proposed, where classical active triangulation based on a single laser stripe is combined with a passive technique for motion estimation. System operation is performed with the device shown in Fig. 1, composed of a laser illuminator and an off-the-shelf camera kept in a fixed relative position. The device is moved manually in front of the object to reconstruct. As with the other hybrid active/passive approaches mentioned above, ours combines good accuracy and flexibility of use. Note that, unlike other hand-held acquisition systems, both the camera and the laser are moved during the scanning. Motion estimation is carried out with two different approaches. The first one is based on Simultaneous Localization and Mapping (SLAM). The second one relies on tracking of a checkerboard pattern in the scene and planar homography decomposition.

In the next Section, a general description of the approach is given. Then Sect. 3 and 4 discuss respectively the SLAM and the homography-based motion estimation approaches. In Sect. 5, a comparison between the motion estimation strategies is addressed, and experimental results are given. Finally, conclusions and directions for future work are discussed in Sect. 6.

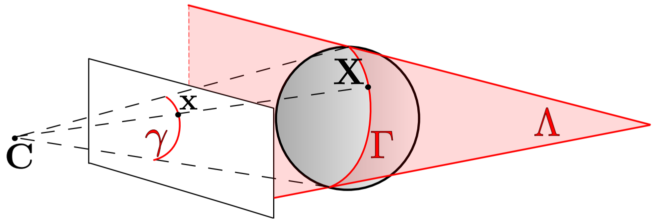


Fig. 2. An example of the device layout while scanning an object. C is the camera center, Λ the laser plane, Γ the 3D laser trace and γ the 2D laser image.

2 The Approach

Figure 2 shows the basic projection geometry of laser profile Γ onto the image. At any time t, each point $\mathbf{x}$ of the imaged laser profile γ can be backprojected onto the laser plane Λ, thus obtaining its pre-image $\mathbf{X} \in \Gamma$. The backprojection equation can be expressed as

$$ {}^{c}\tilde{\mathbf{X}} = \frac{d_\Lambda}{\mathbf{n}_\Lambda^\top \mathtt{K}^{-1}\mathbf{x}} \mathtt{K}^{-1}\mathbf{x} \;, \tag{1} $$

where $\mathbf{n}_\Lambda^\top {}^{c}\tilde{\mathbf{X}} - d_\Lambda = 0$ is the laser plane equation in inhomogeneous camera-centered coordinates, $\mathbf{x}$ is a homogeneous 3-vector, and $\mathtt{K}$ is the camera calibration matrix.

In our approach, object reconstruction is achieved by collating all laser profiles into a single 3D model. This is done by expressing all the backprojected laser profiles in a unique, world-centered reference frame, as

$$ {}^{w}\tilde{\mathbf{X}} = \mathtt{R}_t^\top \left[{}^{c}\tilde{\mathbf{X}} - \mathbf{t}_t \right] \;, \tag{2} $$

where $\{\mathtt{R}_t, \mathbf{t}_t\}$ is the roto-traslation of the camera w.r.t. the world reference frame at time t.

Note that the backprojection map of Eq. (1) does not change with time, since the camera and the laser are in a fixed relative position. Conversely, the camera-world coordinate transformation of Eq. (2) is time-dependent, and must be re-estimated for collating each new laser profile. The approach can be decomposed into four main phases as follows.

1. *System Calibration.* This phase is aimed at estimating the matrix $\mathtt{K}$ and the laser plane parameters to be used in Eq. (1) for the purpose of model building. In this phase, a planar checkerboard pattern is moved by hand in front of the camera-laser system, which is kept in a fixed position. Camera calibration is carried out with the method presented in [14]. For the purpose of laser calibration, a reference frame is attached to the planar calibration pattern, referred to as π, with the normal of the plane coincident with the Z axis, so that the plane equation is simply ${}^{\pi}Z = 0$. For each frame, the

roto-translation $\{\mathtt{Q}, \mathbf{b}\}$ such that ${}^{\pi}\tilde{\mathbf{X}} = \mathtt{Q}^{\top}[{}^{c}\tilde{\mathbf{X}} - \mathbf{b}]$ can be easily obtained as a sub-product of camera calibration [14]. The time-varying pattern plane parameters $(\mathbf{n}_{\pi}, d_{\pi})$ can be then computed as

$$\mathbf{n}_{\pi} = \mathbf{q}_3 \qquad d_{\pi} = \mathbf{q}_3^{\top}\mathbf{b} \ , \tag{3}$$

where $\mathbf{q}_3$ is the third column of $\mathtt{Q}$. Once the pattern plane parameters are computed for all the calibration frames, the set of back-projected laser points $\{{}^{c}\tilde{\mathbf{X}}_i \in \Lambda \cap \pi\}_{i=1}^{N}$ can be computed using Eq. (1) with $(\mathbf{n}_{\pi}, d_{\pi})$ in the place of $(\mathbf{n}_{\Lambda}, d_{\Lambda})$. Since the set contains (thanks to the different planar pattern orientations) at least three non-aligned points, the laser plane parameters are estimated by solving an over-constrained linear system.

2. *Laser Extraction* During the reconstruction process the laser profile is automatically extracted from each image. Given the acquisition video we start by tracking the scanned object using an implementation of the Adaptive Mean Shift [15] in the HSV color space, so to isolate a region of interest (ROI) in each video frame. Then the laser search is performed in a color subspace that depends on the real laser color. For example, using a red laser profiler, only the red channel of the image is used during the extraction. The resulting image is converted into gray levels and the highest intensity pixels are chosen as laser point candidates. Then the gray image is convolved with a Sobel filter to enhance the laser stripe edges. Starting from the candidate pixels, we search the left and right edges of the laser on the filtered image and only those points surrounded by both edges are kept. Finally a Center of Mass algorithm [16] is used to achieve subpixel accuracy.
3. *Model Building.* In this phase the 3D laser profiles are obtained using Eq. (1). Then, the roto-traslation $\{\mathtt{R}_t, \mathbf{t}_t\}$ between the world and camera frames is needed to collate each single 3D profile in an unique reference frame, using Eq. (2). To estimate the camera-laser group movements, two strategies have been implemented and tested. In the first case (see Sect. 3) SLAM algorithm is used to recover the roto-traslation at any time. Alternatively, a homography-based motion estimation algorithm, based on the tracking of a planar pattern, is used in the second case (see Sect. 4).
4. *Texture Acquisition* To augment the raw 3D shape model, the texture of the object is recovered by simply projecting each 3D point back onto the image plane, and then sampling color at the nearest pixel.

3 SLAM Motion Estimation

SLAM algorithms [17, 18] are mainly used to estimate the camera position in an unknown environment. Building an incremental 3D representation of the scene (referred to as map), these algorithms are capable to estimate the camera roto-traslation exploiting the known 2D-3D correspondences between the images and the world. In this work a keyframe-based iterative approach, similar to an incremental structure and motion algorithm, was used. Fig. 3 shows its main steps.

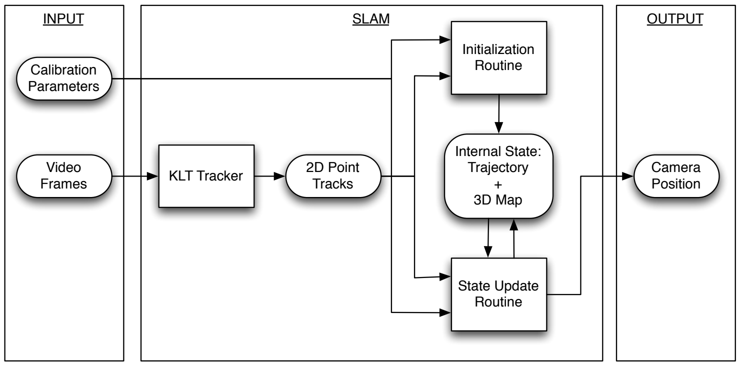


Fig. 3. A block representation of the implemented SLAM algorithm.

Between each subsequent video frame a set of 2D matches is obtained, using an implementation of the KLT tracker [19, 20] and, by linking subsequent matches, the system is able to group the movements of each 2D point in *tracks* — e.g., if A matches with B and B matches with C, than {A,B,C} defines a track for a single point movements.

Given the calibration matrix $\mathtt{K}$, to estimate the first two camera positions and to obtain an initial map representation an *initialization routine* is needed. To achieve this, the world reference frame is attached to the first camera position and a second frame, with sufficient baseline, is manually chosen. After robustly estimating the essential matrix $\mathtt{E}$ from 2D correspondences and decomposing it in $\{\mathtt{R}_1, \mathbf{t}_1\}$ [21], the first two camera matrices are given as $\mathtt{P}_0 = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}_1 = \mathtt{K}[\mathtt{R}_1 \mid \mathbf{t}_1]$. Note that to correctly initialize the real scene scale factor and set the magnitude of $\mathbf{t}_1$, a metric reference (here a checkerboard pattern) has to be visible in the first frames of the sequence. The 3D map is computed by triangulation, storing in a look-up table the correspondence between a 2D track and a 3D point. Finally a bundle adjustment [22] optimization is performed. In this way the system internal state, defined as the camera trajectory and the 3D map, is initialized.

As time progresses, the internal state grows through a *state update routine.* By exploiting the 2D-3D correspondences — easily computed using the updated 2D tracks and the look-up table defined above — the camera matrix at time t is obtained with a robust implementation of a pose estimation algorithm.

To increase the map and to minimize the estimation global error, at specific times a frame is chosen as new keyframe. In this case, in addition to the camera matrix estimation, a new triangulation step is performed, so as to add new 3D points in the map. All the parameters (the keyframe's camera matrices and the map) are then further optimized with a new bundle adjustment iteration.

4 Homography-based Motion Estimation

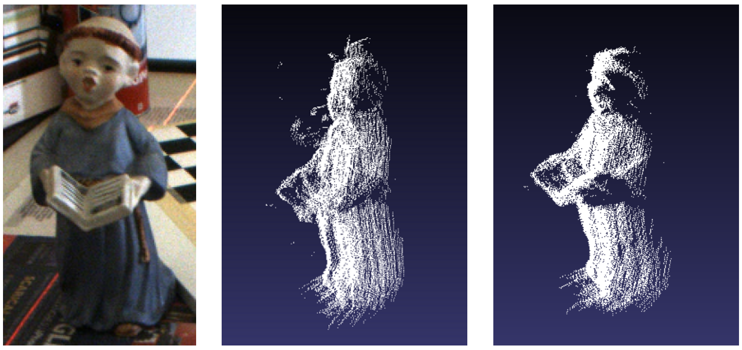
In this case, while the camera-laser device is moved to scan the object, a planar pattern has to be kept still and in view of the camera. This allow us to estimate, as time goes by, the homography $\mathtt{H}_\pi$ between the (moving) image plane and the (fixed) pattern, and eventually compute the roto-traslation $\{\mathtt{R}_t, \mathbf{t}_t\}$ used in Eq. (2) for collating profiles. This is done as follows. The homography has the form

$$\mathtt{H}_\pi = \mu \mathtt{K} \left[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}_t \right] \quad , \tag{4}$$

where $\mathtt{R}_t = [\mathbf{r}_1 \ \ \mathbf{r}_2 \ \ \mathbf{r}_3]$ and μ is an unknown scale factor. Defined $\mathtt{H}_\pi = [\mathbf{h}_1 \ \ \mathbf{h}_2 \ \ \mathbf{h}_3]$, it holds $\mathbf{r}_i = \frac{1}{\mu}\mathtt{K}^{-1}\mathbf{h}_i$ for $i = 1, 2$ and $\mathbf{t}_t = \frac{1}{\mu}\mathtt{K}^{-1}\mathbf{h}_3$. Now, for the orthonormality of $\mathtt{R}_t$, the scale factor and the last column of the rotation matrix can be respectively computed as $\mu = \|\mathtt{K}^{-1}\mathbf{h}_1\|$ and $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$.

5 Results

A comparison of the results obtained with the two motion estimation solutions is given hereafter. Operationally speaking, SLAM is preferable, as it guarantees a higher flexibility in terms of objects size and choice of viewpoint. In fact the homography-based solution is more constrained in this sense since users must take care that the checkerboard patter always remain in view during the acquisition. However, as evident from the qualitative result of Fig. 4, homography-based approach outperforms SLAM for what concerns 3D model accuracy. Indeed, as



(a) (b) (c)

Fig. 4. An example of reconstruction obtained with the SLAM (Fig. 4(b)) and the Homography-based (Fig. 4(c)) solutions.

it's possible to see in Fig. 5, the SLAM and the homography-based computed trajectories show an increasing divergence. Motion estimates are initially very similar but then SLAM performance gradually degrades. The main reason of

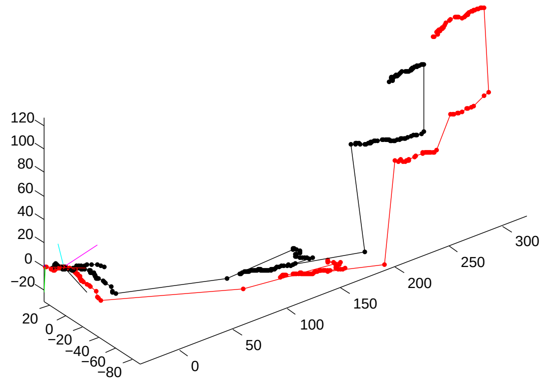


Fig. 5. Results of camera trajectory estimation for the reconstruction of Fig. 4. In red the trajectory computed with the SLAM algorithm. In black the homography-based estimated trajectory. As we can see, while the estimation goes on, the differences between the two trajectories increase. Note that the points on the trajectories are the camera center positions. Instead the straight lines where no points are drawn represent sub-sequences where no laser stripe was found over the scanned object and so no motion estimation was carried out.

this behavior is to be found in the scale factor drift that generally affects single camera structure and motion algorithms.

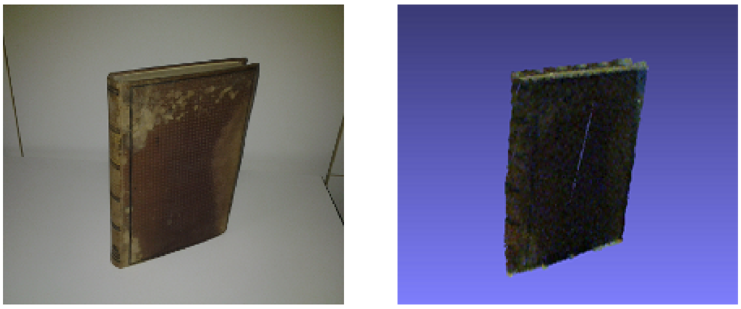
More reconstruction tests were carried out with the homography-based approach. Figure 6 shows the results of the 3D reconstruction algorithm for other two different objects. To evaluate the accuracy of the reconstructions, we compared the 3D models with several measurements manually done on the real objects with an high precision caliber. Table 1 reports on the measurements and errors for the *Book* object of Fig. 6(a). Table 2 summarizes the accuracy result in terms of average and maximum error. The results are good and comparable to other approaches requiring either more constrained acquisition scenarios or more sophisticated hardware. On the other hand, a closer inspection to Fig. 6 reveals that the acquired 3D models present some gaps, that are mainly due to the fact that manual operation does not guarantee that all surface points are illuminated at least once by the laser stripe.

Table 1. Measurements (cm) of the *Book* model.

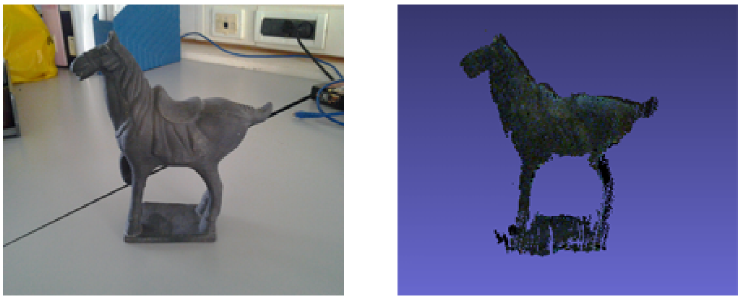
Dim.	Real	Model	Error
Height	22.50	22.13	0.37
Width	14,80	14,71	0.09
Length	2.08	2.38	0.30

Table 2. Average and max errors (cm) for the *Book* and *Horse* models.

Model	Avg. Error	Max Error
Book	0.25	0.37
Horse	0.23	0.4



(a) Example 1: Book



(b) Example 2: Horse

Fig. 6. Reconstruction examples: in the first column the photos, in the second the 3D models.

6 Conclusions and Future Work

In this work we described a tool based on an active/passive framework for reconstruction of realistic 3D models of limited size objects, and we compared two motion estimation solutions. System operation includes a simultaneous camera and laser calibration phase, followed by backprojection and collation of all imaged laser profiles. As shown in Sect. 5, the implemented (mono) SLAM solution, which has been found to be perfectly adequate for augmented reality applications (see Fig. 7 and its description), appears to be less accurate than the homography-based approach for 3D reconstruction applications. On the other hand, the homography-based approach is less flexible than SLAM, but nevertheless allows us to obtain good quality models using a very simple procedure and an inexpensive hardware. The software is currently implemented as a prototype with partial manual operation — as for example the checkerboard detection step.

Future work will address code optimization — including the removal of all manual operations — and the development of suitable point cloud densification strategies aimed at filling the model gaps. In addition a new device with a stereo pair and a laser emitter is currently under study. Using a stereo approach

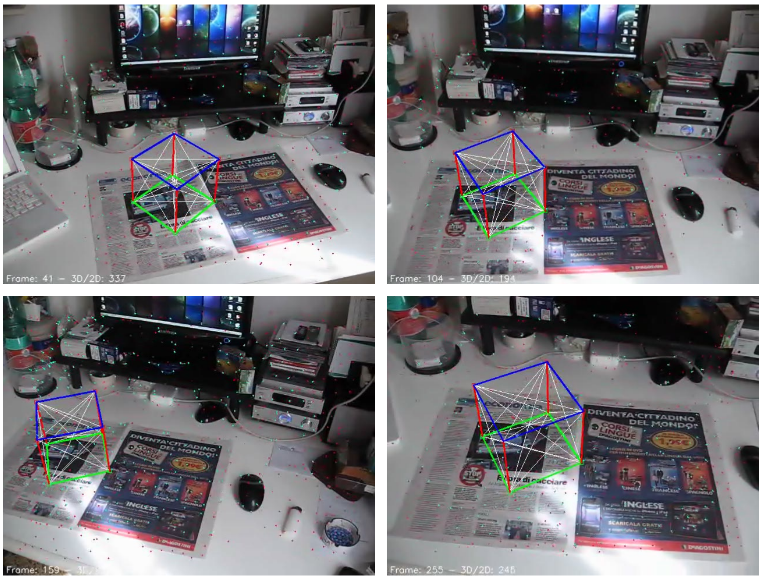


Fig. 7. An augmented reality (AR) application using our SLAM algorithm. The virtual wire-frame cube undergoes the correct perspective deformations and remains stable upon the desk. Although the algorithm is the same used for 3D reconstruction, here the perceived quality of motion estimation is higher. In fact, differently from Fig. 4(b), the inaccuracies in camera motion estimates do not appear as flaws. This shows that 3D reconstruction requires a higher estimation accuracy than AR to achieve a similar perceptual quality.

for SLAM is likely to lead to a more robust motion estimation, avoiding any scale factor uncertainty, and yield an even more efficient an flexible tool for 3D structure recovery.

Acknowledgements

This work has been carried out during the THESAURUS project, founded by Regione Toscana (Italy) in the framework of the “FAS” program 2007-2013 under Deliberation CIPE (Italian government) 166/2007.

References

1. Chen, F., Brown, G.M., Song, M.: Overview of Three-Dimensional Shape Measurement using Optical Methods. Optical Engineering **39** (2000) 10–22
2. Bernardini, F., Rushmeier, H.E.: The 3D Model Acquisition Pipeline. Computer Graphics Forum **21** (2002) 149–172
3. Rocchini, C., Cignoni, P., Montani, C., Scopigno, R.: A low cost 3D scanner based on structured light. Computer Graphics Forum **20** (2001) 299–308

4. Winkelbach, S., Molkenstruck, S., Wahl, F.M.: Low-cost laser range scanner and fast surface registration approach. In: Proceedings of the 28th conference on Pattern Recognition. DAGM'06, Berlin, Heidelberg, Springer-Verlag (2006) 718–728
5. Habbecke, M., Kobbelt, L.: Laser brush: a flexible device for 3d reconstruction of indoor scenes. In: Proceedings of the 2008 ACM symposium on Solid and physical modeling, New York, NY, USA, ACM (2008) 231–239
6. Colombo, C., Comanducci, D., Del Bimbo, A.: Shape reconstruction and texture sampling by active rectification and virtual view synthesis. Computer Vision and Image Understanding **115** (2011) 161–176
7. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a Day. In: Proceedings of the International Conference on Computer Vision. ICCV '09, Kyoto, Japan (2009)
8. Farenzena, A.M., Fusiello, A., Gherardi, R.: Structure-and-Motion Pipeline on a Hierarchical Cluster Tree. In: Proceedings of the IEEE International Workshop on 3-D Digital Imaging and Modeling, Kyoto, Japan (2009)
9. Vogiatzis, G., Hernàndez, C.: Video-based, real-time multi view stereo. Image and Vision Computing (2011)
10. Wang, L., Liao, M., Gong, M., Yang, R., Nistèr, D.: High-quality real-time stereo using adaptive cost aggregation and dynamic programming. In: 3rd Int. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), Springer-Verlag (2006) 798–805
11. Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M.: Shape from shading: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(8) (1999) 690–706
12. Bouguet, J.Y., Perona, P.: 3D Photography Using Shadows in Dual-Space Geometry. International Journal of Computer Vision (IJCV) **35** (1999) 129–149
13. Hernàndez, C., Vogiatzis, G., Cipolla, R.: Multi-view Photometric Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008)
14. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 1330–1334
15. Bradski, G.R.: Computer Vision Face Tracking For Use in a Perceptual User Interface. Intel Technology Journal (1998)
16. Fisher, R.B., Naidu, D.K.: A Comparison of Algorithms for Subpixel Peak Detection. In: Image Technology, Advances in Image Processing, Multimedia and Machine Vision, Springer-Verlag (1996) 385–404
17. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07), Nara, Japan (November 2007)
18. Mei, C., Sibley, G., Cummins, M., Newman, P., Reid, I.: RSLAM: A system for large-scale mapping in constant-time using stereo. International Journal of Computer Vision (2010) 1–17 Special issue of BMVC.
19. Shi, J., Tomasi, C.: Good features to track. Technical report, Ithaca, NY, USA (1993)
20. Bouguet, J.Y.: Pyramidal implementation of the Lucas-Kanade feature tracker description of the algorithm (2000)
21. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, ISBN: 0521540518 (2004)
22. Lourakis, M.I.A., Argyros, A.A.: SBA: a software package for generic sparse bundle adjustment. ACM Transactions on Mathematical Software (2009) 1–30