AUTOMATIC 3D SURFACE RECONSTRUCTION BY COMBINING STEREOVISION WITH THE SLIT-SCANNER APPROACH

A. Prokos[1], G. Karras[1], E. Petsa[2]
[1] Department of Surveying, National Technical University of Athens (NTUA), GR-15780 Athens, Greece
[2] Department of Surveying, Technological Educational Institute of Athens (TEI-A), GR-12210 Athens, Greece
e-mail: anthpro@central.ntua.gr, gkarras@central.ntua.gr, petsa@teiath.gr

KEY WORDS: photogrammetric scanning, surface reconstruction, triangulation, epipolar geometry, camera calibration

ABSTRACT:

In this paper, a 3D surface scanner is presented. Combining stereovision and slit-scanning, our system is composed of two cameras and a hand-held laser plane. The camera pair is calibrated using synchronized image pairs of a coded chessboard; its imaged nodes are automatically identified, referred to the object points and introduced into a self-calibrating bundle adjustment. For the scanning process, stereoscopic profiles are continuously recorded as the 3D surface is swept by the laser line. After epipolar resampling of the synchronized image pairs, search for point correspondences is thus reduced to identifying intersections of image rows with the recorded laser profiles. The maxima of Gaussian curves fitted to the gray-value data along the epipolar image rows provide initial estimates for peak positions, which are then refined using information from their neighbourhood. In our setup, 3D reconstruction by simple stereovision is strengthened by enforcing extra geometric constraints. First, the coplanarity constraint is imposed on all 3D points reconstructed from a single laser stripe, and the coefficients of all laser planes participate as unknowns in the 3D reconstruction adjustment. Additionally, this also allows identifying mismatches since epipolar lines may have more than one peaks; the correct 3D point is established according to a distance threshold from the laser plane. The solution is further reinforced by placing the object in a corner formed by two background planes (which are scanned along with the object), whose coefficients are also unknowns in the 3D reconstruction adjustment. The linear laser segments produced on either side of the object have to satisfy the equation of both the corresponding plane and the laser plane. Image pairs of the corner without the object (longer laser segments) are added to the dataset for a more accurate determination of plane equations. Results are presented and evaluated from this setup, whose typical accuracy is estimated in the order of 0.2 mm in 3D depth estimation.

1. INTRODUCTION

Recent years are witness to a growing demand for 3D surface models in several fields (e.g. cultural heritage documentation or industrial metrology). Ideally, the 3D models must be generated rapidly and accurately by automatic techniques. As a response to this demand, a number of image-based scanners, both commercial and low-cost ones, have been reported (Forest & Salvi, 2002; Blais, 2004). Of course, stereovision remains a standard approach. Its main problem is finding point correspondences, in particular when dealing with surfaces of low texture.

A way to overcome this problem is replacing the second camera by devices which project various patterns (e.g. structured light) but can be as simple as a laser plane. Most common among such triangulation-based 'range-finders' are those using laser planes (i.e. projection of laser stripes), also referred to a 'slit-scanners'. Such systems typically combine a camera and a projected laser plane which intersects the object surface to highlight a profile. The 3D points of each profile are found (without redundancy) as intersections of the laser plane and the projection rays defined by the respective image points of the profile.

Thanks to its simplicity, several low-cost systems of the 'slit-scanner' type have been reported. If the laser plane is moved by hand independently from the camera, its position in space must be calculated for each image. In Zagorchev & Goshtasby (2006) this is achieved through the intersection of the laser plane with a reference double-frame, whereas Winkelbach et al. (2006) use two external orthogonal planes intersected by the laser plane. A scanner with simpler components is that of Bouguet & Perona (1998) using one camera and the shadow produced form a hand-held moving rod. In a very interesting implementation, Kawasaki & Furukawa (2007) use the mere fact that laser lines define coplanar object profiles (implicit coplanarity) to acquire dense 3D data, while also exploiting coplanarity information from object planes in the scene (explicit coplanarity). Projective results are upgraded to Euclidean via suitable constraints, e.g. orthogonal planes (if necessary, the constraints are supplied by a device producing two orthogonal laser planes).

Following Prokos et al. (2009), this paper presents a low-cost photogrammetric range-finder which combines stereovision and the slit-scanner principle. The two web cameras are automatically calibrated to provide both their interior orientations and their true to scale relative orientation. A laser plane generator is used to code the scene and, hence, simplify the correspondence problem to a peak detection question. Our setup is, essentially, similar to that of Davis & Chen (2001), yet the straightforward solution from stereovision is here enhanced by additional geometric constraints. The fact that all points on a single laser profile belong to the same plane (laser plane) is exploited to pose a coplanarity constraint to these points. Furthermore, the 3D object is placed in the corner formed by two unknown intersecting planes; thus, the end-segments of the depicted laser profiles (intersections s_1, s_2 with the background planes p_1, p_2 in Fig. 1) are straight. This is exploited to introduce a further coplanarity constraint as each such segment also belongs to the corresponding background plane. The equations of these planes are estimated in the reconstruction algorithm (in which, along with the image pairs for surface scanning, images of the planes without the object may also participate).

It is to note that this (optional) second coplanarity constraint is not always used, since the linear profile end-segments may not be sufficiently long to provide reliable data if larger image parts are occupied by the 3D object (this allows higher resolution in object space).

2. SYSTEM DESCRIPTION

The system consists of a pair of web cameras in fixed relative position and a hand-held laser stripe generator. The 3D object is placed close to the intersection of two (unknown) background planes. Thus, the hardware components of the system are:

- Two 640x480 colour web cameras, fixed in a constant relative position throughout the process. The system is calibrated automatically as explained below. In a typical application, the mean pixel size in object space is ~0.8 mm.
- A green line laser with adjustable focus, allowing width of the laser line ~0.5 mm.
- A typical black-and-white chessboard pattern (with one red square to fix the object system) for calibration purposes.
- Two planes p_1, p_2 forming a corner (optional).

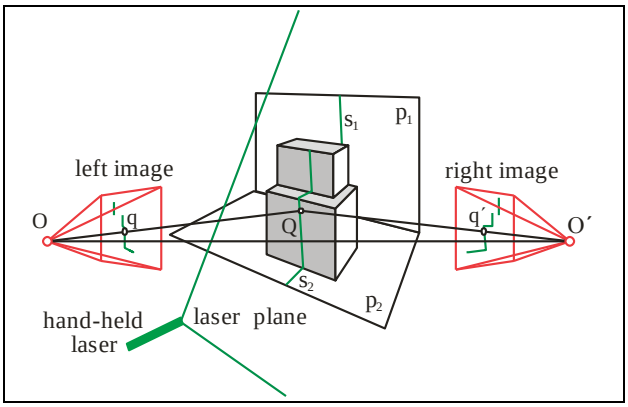

Figure 1. The system setup

As the 3D object surface is manually 'swept over' with the laser stripe, the cameras capture synchronized frames of the scene recording the object, the background planes and the laser stripe which 'codes' the scene. If homologous points q, q′ of the laser profile are identified, the projective rays thereby defined intersect at the 3D point Q (Fig. 1). In addition, all 3D points resulting from a single laser profile are bound to be coplanar (on the laser plane). Furthermore, all 3D points reconstructed from laser profiles on a background plane must simultaneously belong to yet another plane (the corresponding background plane). These two constraints introduce a significant redundancy in the adjustment, thereby allowing higher accuracy and reliability.

3. THE SCANNING PROCESS

3.1 System calibration

Our group has presented an algorithm* which accepts a number of images of simple planar chessboard patterns to automatically estimate the interior orientation of the camera used (Douskos et al., 2008). The algorithm first extracts the chessboard nodes via a Harris operator, orders them and finally determines the camera geometry elements by bundle adjustment. Since here the scaled relative orientation of the cameras is also required, input to our modified calibration algorithm is synchronized image pairs of a chessboard pattern of known grid size and with one of its black squares changed to red. The latter is automatically detected, and thus the origin of the chessboard coordinate system can be fixed (see Prokos et al., 2009, for more details). Evidently, the collinearity equation used for the left camera

$$\begin{bmatrix} x - x_{o1} \\ y - y_{o1} \\ -c_1 \end{bmatrix} = \lambda \cdot \boldsymbol{R}_1 \begin{bmatrix} X - X_{o1} \\ Y - Y_{o1} \\ Z - Z_{o1} \end{bmatrix}$$

is modified for the right camera to accommodate the matrix of relative rotations $\boldsymbol{R}_{12}$ and the three base components:

$$\begin{bmatrix} x - x_{o2} \\ y - y_{o2} \\ -c_2 \end{bmatrix} = \lambda \cdot \boldsymbol{R}_{12} \cdot \boldsymbol{R}_1 \cdot \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \left(\begin{bmatrix} X_{o1} \\ Y_{o1} \\ Z_{o1} \end{bmatrix} + \boldsymbol{R}_1^T \cdot \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} \right) \right)$$

In all calibration adjustments performed, the standard error was about 0.2 pixels. A typical calibration output is seen in Table 1 (k_i, p_i are the coefficients of lens distortion). In Fig. 2 a typical stereo pair used for calibration is shown.

Table 1.
Calibration results (from 20 image pairs)

σ_o = 0.21pixel		
	left camera	right camera
c_x (pix)	957.51 ± 0.32	957.83 ± 0.25
c_y (pix)	954.11 ± 0.33	958.09 ± 0.28
x_o (pix)	23.31 ± 0.55	-67.14 ± 0.73
y_o (pix)	-7.14 ± 0.39	-9.70 ± 0.51
$k_1(\times 10^{-07})$	-1.39 ± 0.03	-1.65 ± 0.02
$k_2(\times 10^{-13})$	-1.05 ± 0.17	-7.12 ± 0.12
$p_1(\times 10^{-06})$	-3.26 ± 0.22	-1.60 ± 0.28
$p_2(\times 10^{-06})$	-4.18 ± 0.16	-1.59 ± 0.19
	relative orientation	
Bx (cm)	38.86 ± 0.01	
By (cm)	2.00 ± 0.00	
Bz (cm)	-9.16 ± 0.02	
ω (°)	-5.91 ± 0.04	
φ (°)	31.46 ± 0.05	
κ (°)	12.85 ± 0.01	

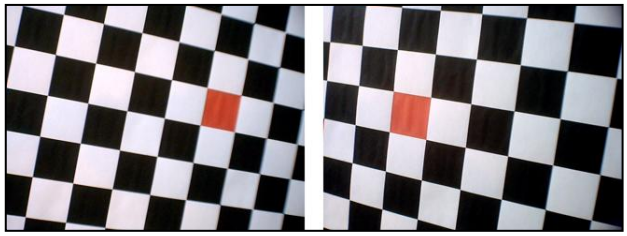

Figure 2. A stereo pair used for the calibration.

3.2 Image acquisition and subtraction

For scanning, stereo pairs are continuously taken from each position of the static camera system; each pair records the instantaneous profile of the 3D surface which is intersected by the laser plane as the latter is slowly moved manually over the surface.

Dull surfaces may be scanned with normal illumination of the scene (which may also be sufficient for capturing object texture of good quality); shiny surfaces need to be scanned with no exterior light source. Either way, laser profiles have to be isolated from the background, i.e. from all frames a reference image (generated here as the temporal median of a few images) has to be subtracted. If illumination is good, a further use of these background images is to supply all surface points with their specific photo-texture for the purposes of visualization; else, from each scanning position an extra image pair may be taken under suitable illumination simply for photo-texturing.

* The source code in Matlab of the calibration toolbox FAUCCAL, with documentation, tips and test imagery, is available on the Internet at: *http://www.survey.ntua.gr/main/labs/photo/staff/gkarras/fauccal.html*

3.3 Peak detection on epipolar images

Using the calibration data all image pairs are transformed to epipolar pairs, whereby known systematic images errors (here lens distortion) are removed. Thus, the search for homologous points on the laser profile is confined on corresponding epipolar lines (image rows), i.e. peaks must be determined on each image row. Several peak detection approaches have been reported (Fisher & Naidu, 1996). Here, a Gaussian curve is adapted directly to the intensity values of each row:

$$f(x) = ae^{-\frac{(x-b)^2}{2\sigma^2}} + d$$

First, a threshold is applied to each row, providing an estimation of the position of the peak (or peaks) on an epipolar line as well as peak width. Curves are fitted only to positions which yielded widths below a limit, in order to exclude 'stretched' stripes due to laser planes intersecting the 3D surface at a small angle. The subpixel estimation of the peak position is given by parameter b in the above equation.

However, this peak estimation uses data only in the direction of the image row. Thus, in order to relax the strictness of 1D interpolation, two additional Gaussian curves with a common b parameter in the image x-direction are simultaneously fitted, namely in the directions of the two main diagonals through the initial peak estimation. The data from the two diagonals contribute in the estimation with smaller weight. Consequently, the final peak position remains on the epipolar line, but is influenced by gray values from the neighborhood of the initial estimation. In Fig. 3 one may see the effect of this procedure.

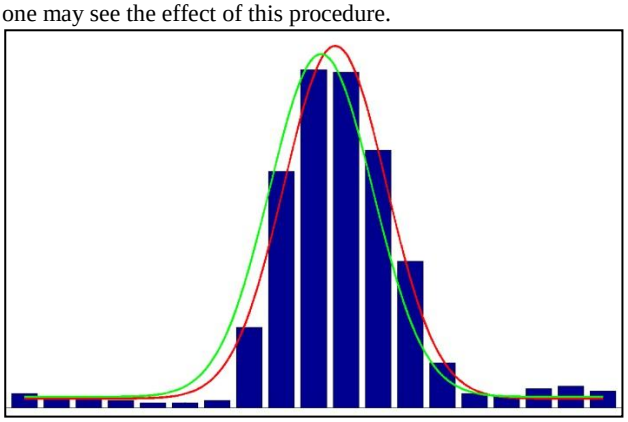

Figure 3. Profile along an epipolar line. Curve fitted only along the image row (red) and curve fitted together with curves along the main diagonals (green).

Before extracting peak positions, a 3x3 Gaussian filter removes image noise. This mild filter was generally sufficient, due to the good quality of the employed laser (for lasers of poorer quality used in previous experiments a median filter had to be applied first). It is pointed out that an increase of the window size of the filters may give better precision in peak detection; however, the end result of the 3D point cloud will probably be too smooth. For lines with multiple peak encounters (e.g. close to occlusion borders or due to reflections) the peaks are stored separately and processed as explained later.

3.4 Background planes

Prior to scanning the object, the background planes are scanned. After peaks on epipolar lines have been identified for all point pairs as outlined above, their x_1, x_2 and y image coordinates are used in the simple parallax equations in order to reconstruct the 3D points:

$$X = \frac{Bx_1}{p} \qquad Y = \frac{By}{p} \qquad Z = -\frac{Bc}{p} \qquad (p = x_1 - x_2)$$

The points of the background planes must be separated into two groups, each representing the respective plane. This is done for each image pair by fitting two 3D lines using RANSAC. End result is two point clouds, from which coefficients of the background planes are estimated.

3.5 Reconstruction algorithm

As regards object scanning, very good initial estimations of the 3D position of all points of a laser profile are obtained from the parallax equations; from these the coefficients of the laser plane are estimated. Together with the coefficients of the background planes, this allows sorting profile points in three groups, namely points on the two background planes and object points. But the parallax equations yield 3D object points without redundancy (i.e. without a means for estimating precision or for gross error detection). In our approach the answer to this, as mentioned, is the introduction of extra geometric constraints.

First, triangulation is strengthened by the additional constraint that all 3D points reconstructed from a recorded laser stripe are coplanar. Thus, the coefficients of all laser planes are involved as unknown parameters in the adjustment. A further constraint is enforced by means of the two background planes (also intersected by the laser plane). Obviously, the end-parts of the laser profiles on either side of the object are straight (Fig. 4). Therefore, their points must simultaneously satisfy the equations of the corresponding laser plane as well as those of the corresponding background plane. Estimates for the coefficients of the two background planes are known from scanning the corner beforehand.

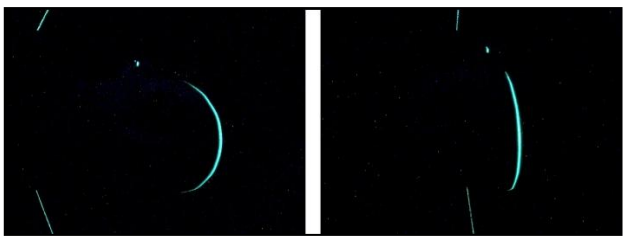

Figure 4. A typical stereo pair used in the scanning process. Besides the object, the laser plane intersects the two background planes producing linear segments on either side of the object.

In Prokos et al. (2009) each laser profile was adjusted independently, i.e. the parallax equations were combined with the laser plane equation and soft constraints for the linear end-segments. Consequently, in each adjustment a total of 2×N + 3 unknowns were involved, namely the X and Z coordinates of all N points of the laser profile (Y-values are directly found afterwards from the final p-values) and the laser plane coefficients. Here, on the contrary, the robust approach of a unified 3D reconstruction adjustment has been adopted. This means that all laser profiles recorded from a particular viewpoint of the camera system are adjusted together, with individual points forced to belong to their (unknown) laser plane and, if they are points of end-segments, also constrained to lie on the corresponding (unknown) background plane. Thus, unknowns here are the X and Z coordinates of all points of the n profiles plus 3×n coefficients of the laser planes plus 6 coefficients of the background planes (which are the common unknowns). It is noted that, in order to have longer linear end-segments and also include observations close to the

intersection of the planes, the images of the background planes without the object are also included in the fitting adjustment.
The second constraint is optional, in the sense that the object might not be placed in a corner or, if placed, background planes might be only marginally visible in the images to allow the object to occupy the largest possible image part (higher resolution in object space). In such a case, an overall adjustment is clearly pointless, i.e. each profile is processed independently.

An extra step is to back-project all 3D points onto the pair of reference images in order to interpolate sets of RGB values which complement the 3D data to produce a final XYZ–RGB set. Finally, the results from the different scanning sessions (from the different viewpoints of the camera systems) are co-registered in a single 3D surface model using ICP.

3.6 Multiple peaks

Epipolar lines which produce more than one peak are stored separately and do not participate in the solution, i.e. reconstructed are at first only points resulting from epipolar lines with a single peak. After the adjustment, 3D points are calculated for all possible combinations of stored multiple peaks on epipolar lines. The actual object points among them are separated from the outliers by means of a distance threshold from the estimated laser plane (Prokos et al., 2009). This is a further exploitation of the fact that all points of a profile are coplanar.

4. APPLICATION AND EVALUATION

4.1 Expected accuracy

The precision of 3D coordinates is directly related to the error σ_p of the x-parallax (p), which is the result of the uncertainty σ_x in the x-direction of peak positions estimated through Gaussian curve fitting (parameter *b*). The parallax error is propagated in 3D space through the image scale and the base-to-distance ratio. For the setup of Table 1 (c = 950 pixel, B = 40 cm), an average imaging distance of 70 cm in scanning the test cylinder (see below) and $\sigma_x = \pm 0.1$ pixel for the uncertainty of peak estimation (i.e. $\sigma_p = \pm 0.15$ pixel), the typical expected precision in depth is estimated as $\sigma_z = \pm 0.2$ mm.

4.2 Evaluation of accuracy

The validity of the above estimation was checked by scanning a white PVC plumbing tube with a nominal diameter of 125 mm. A cylinder was fitted to the 3670 XYZ values of the point cloud from one scanning position which represented approximately 2/5 of the perimeter. The standard error of the surface-fitting adjustment was 0.2mm (the same as in Prokos et al., 2009).

4.3 Practical applications

Objects scanned with our system were a polyester souvenir statue of Venus (height ~15cm), a 1985 Australian dollar coin and the face of one of the authors. The first object was scanned with a 20 cm base; the 3D model is seen in Fig. 5 (top). The coin is a rather extreme case, since the system has not been designed for very small objects. A 10 cm base was used. Crucial was here the width of the laser stripe: pixel size was less than 0.1mm, but the laser line could not be narrower than 0.5mm, i.e. 5 pixels. The end result, shown in Fig. 5 (middle), was noisy but the surface appears to be adequately captured. The last object, scanned with a 30 cm base from two viewpoints, also represents an extreme case since the person should remain frozen during the scanning phase which, at the moment, takes several minutes. Under these circumstances, the result (seen in Fig. 5, bottom) is satisfactory. The images used to drape this 3D model with texture where not created with the temporal median approach but taken separately, since the 3D point cloud was acquired with the subject's eyes closed.

Figure 5. Images and final 3D models: small statue of Venus (top), coin (middle) and face of the primary author (bottom).

5. CONCLUSION

An implementation of a low-cost 3D scanner has been reported, based on the combination of the stereovision and the slit scanner principles, accompanied by the introduction of extra geometric constraints. Compared to previous work (Prokos et al., 2009), a main goal here was to improve the overall reconstruction reliability. This has been achieved by the unified adjustment of all laser profiles from each scanning position. Some computational problems have to be solved if all laser profiles from all scanning viewpoints are to be adjusted in a single solution. Future tasks include establishing further means for detecting outliers within but also between point clouds from different scanning positions.

REFERENCES

Blais F., 2004. Review of 20 years of range sensor development. Journal of Electronic Imaging, 13(1), pp. 231-240.

Bouguet J.-Y., Perona P., 1998. 3D photography on your desk. Proc. IEEE Int. Conf. on Computer Vision, pp. 43-50.

Davis, J., Chen, X., 2001. A laser range scanner designed for minimum calibration complexity. Proceedings of Third International Conference on 3-D Digital Imaging and Modeling. pp. 91- 98

Douskos V., Kalisperakis I., Karras G.E., Petsa E., 2008. Fully automatic camera calibration using regular planar patterns. Int. Arch. Phot. Rem. Sens., 37(B5), pp. 21-26.

Fisher R.B., Naidu D.K., 1996. A comparison of algorithms for subpixel peak detection. Advances in Image Processing, Multimedia and Machine Vision. Springer, pp. 385-404.

Forest J., Salvi J., 2002. A review of laser scanning three-dimensional digitisers. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 1, pp. 73-78.

Kawasaki H., Furukawa R., 2007. Dense 3D reconstruction method using coplanarities and metric constraints for line laser scanning. 6th IEEE Int. Conf. on 3D Digital Imaging and Modeling (3DIM '07), pp. 149 – 158.

Prokos A., Karras G., Grammatikopoulos L., 2009. Design and evaluation of a photogrammetric 3D surface scanner. Proc. 22nd CIPA Symposium, October 11-15, Kyoto, Japan.

Winkelbach S., Molkenstruck S., Wahl F.M., 2006. Low-cost laser range scanner and fast surface registration approach. Proc. DAGM '06, Lecture Notes in Computer Science, 4174, Springer, pp. 718–728.

Zagorchev L., Goshtasby A.A., 2006. A paint-brush laser range scanner. Computer Vision & Image Understanding, 101, pp. 65-86.