DESIGN AND EVALUATION OF A PHOTOGRAMMETRIC 3D SURFACE SCANNER

A. Prokos[1], G. Karras[1], L. Grammatikopoulos[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, lazaros.pcvg@gmail.com

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

ABSTRACT

This paper presents a low-cost 3D surface scanner, composed of two fixed web cameras and a hand-held planar laser beam. Setup pre-calibration provides interior orientations of the cameras and their scaled relative orientation. Our calibration algorithm, based on bundle adjustment, uses image pairs of a chessboard, whose nodes are identified automatically and referred to the ground points. For scanning, synchronized image pairs are continuously recorded from each location of the static cameras as the laser source is slowly moved by hand; each pair thus records a profile of the 3D surface intersected by the laser plane. Epipolar resampling reduces the search for point correspondences to finding the intersections of homologous epipolar lines with the recorded laser profile. After a smoothing operation, peaks are identified as the maxima of Gaussian curves fitted to the gray-value data along the epipolar lines; the final identification of peaks involves information from the neighbourhood of the initial estimation. An innovative aspect is that the photogrammetric triangulation of 3D points gains in robustness by enforcing extra geometric constraints. Thus, all points of a profile must lie on a laser plane, whose coefficients are involved as unknowns in the 3D reconstruction adjustment. This allows identifying blunders in peak detection; for epipolar lines with more peaks, only points which, when reconstructed, satisfy a distance threshold from the laser plane participate in the final 3D data set. Furthermore, the object is placed in a corner (the equations of its two planes in the setup system are found automatically by prior scanning), which is intersected by the laser plane in two lines. Their points are identified and constrained to simultaneously satisfy both the corresponding plane equation and the equation of the laser plane. Using available modeling software, individual scans are finally triangulated and merged into a single 3D surface model, which may also be draped with image photo-texture. First results indicate that an accuracy better than 0.3 mm appears as feasible with this setup.

1. INTRODUCTION

Recently, there is a growing demand (also in the field of archaeology or cultural heritage conservation) for 3D surface models, particularly of smaller objects, created by techniques both fast and accurate. A variety of non-contact optical surface digitizers or 'range finders' – including commercial and, to a growing extent, low-cost ones – have been proposed to meet this demand, as reviewed in the literature (Forest & Salvi, 2002; Blais, 2004).

Stereovision is such a method; establishing of correspondences between images for triangulating 3D points, however, is a not an easy task, even under the epipolar constraint (Salvi et al., 2004). This can be overcome by projecting randomly patterned texture onto the object surface (D'Apuzzo, 1998). Alternatively, 'depth from triangulation' is further facilitated if the second camera is replaced by a device which defines planes, or rays, in 3D object space. The coded structured light approach, based on recording a light pattern projected onto the scene, allows locating coded points in the image and easily establishing the correspondences with those of the projected pattern; such correspondences allow, via triangulation, the extraction of 3D data. In Salvi et al. (2004) an overview of existing approaches in this field are presented.

As observed by Blais (2004), the slit scanner – usually based on a camera and a line laser – is probably the most commonly used triangulation-based digitizer, due to its potential for optical and mechanical simplicity as well as low cost. The 3D points are reconstructed as intersections of laser planes and projection rays defined by the laser trace on the object surface as recorded on the image. Many commercial systems rely on this principle, utilizing precise mechanisms to control movement of the laser or the object. Yet, simple and more flexible approaches with hand-held laser beams have also been reported. In this case each laser plane must be oriented on-line in space. This is done by further mechanisms or sensors (Blais, 2004), but simpler means are too feasible based on a priori information about the imaged scene. In Zagorchev & Goshtasby (2006) the stripe scanning the object is moved by the user independently from the camera. The laser beam is calibrated from its intersection points with a reference double-frame which is fixed with respect to the object, thus producing object coordinates in the same system. This approach is further simplified in the system of Winkelbach et al. (2006) for 3D data acquisition and registration. Its hardware components are a standard camera, calibrated through markers on the background, and a hand-held line laser sweeping over the surface. Laser planes are calibrated from the stripes on the images which refer to intersections not only with the unknown object surface but also with a background of known geometry (in practice, two intersecting planes). This allows fixing the laser plane in space; surface points can then be triangulated. Instead of a laser plane, on the other hand, Bouguet & Perona (1998) simply exploit the images of the shadow which a stick casts onto the object surface when illuminated by a fixed desk-lamp (point light source). In their system, calibrated using a planar chess-board, the stick is moved by the user in front of the light source, while the shadow plane is determined from its instantaneous trace on a planar surface under the object; triangulation of 3D surface points is then feasible. In a variation of this approach Fisher et al. (1999) have used the shadows cast by a rod of triangular profile, designed for easy detection and pose estimation from a single view.

The mere fact that line lasers geometrically define coplanar object points is used by Kawasaki & Furukawa (2007) to realize an active measuring system without need for explicit calibration of laser planes or camera calibration targets. This method, using a single camera, estimates dense 3D shape and camera geometry by using coplanarity information in the scene extracted from the

laser traces (implicit coplanarity) or from object planes (explicit coplanarity). In the presence of intersecting laser traces (posing constraints among laser planes) this leads to a projective reconstruction, which is upgraded to Euclidean if metric constraints (e.g. orthogonalities of planes) are available; if necessary, these are provided by a device producing two orthogonal laser planes.

The scanner described in this paper combines stereovision and the 'slit scanner' principle. Two cameras – which are first calibrated automatically – view the trace of a moving hand-held laser plane in the scene. The laser line helps solve the correspondence question for triangulating 3D object points with no redundancy. Triangulation is strengthened by the extra constraint that all reconstructed points of a laser line are coplanar. Further, the end-parts of these image lines are straight, as they represent intersections with two background planes which form a corner behind the object; hence, these segments must also be constrained on the respective planes (which have been reconstructed before object scanning). Introduction of such additional 'implicit' and 'explicit' constraints (to adopt the terms of Kawasaki & Furukawa, 2007) makes the approach more robust. Clearly, the method may also function in a pure stereovision mode, i.e. without the additional constraints, and in particular without the background planes (in the latter case, larger image parts may be covered by the object surface, allowing higher resolution in object space).

2. DESCTIPTION OF THE SYSTEM

The scanner is composed of two web cameras and a hand-held laser plane, which 'sweeps' the object to create successive profiles. The cameras, which are in a known fixed position, record the scanned surface in synchronized image pairs (Fig. 1). Intersection of the two homologous rays, defined by the peaks q, q′ of the laser profile, allows computing the 3D coordinates of the points (Q) of the trace. Adding to this, geometric constraints are imposed in the model: 1) all reconstructed profile points belong to a laser plane; 2) at the same time, the reconstructed 3D points of linear segments s_1 and s_2, created as intersections of the laser plane with background planes p_1 and p_2, must also lie on these planes. Equations of these two planes are estimated by previous scanning and can be used to enforce 'hard' or 'soft' constraints.

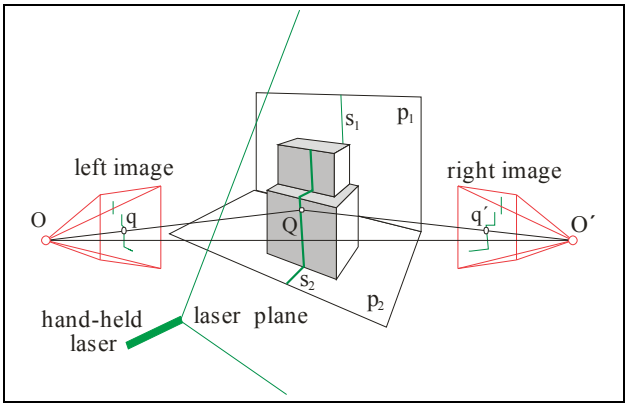

Figure 1. The system setup

The hardware components of the system are:

■ Two 640×480 colour web cameras, fixed on an aluminum bar to keep them in a constant relative position throughout the process. Camera calibration and relative orientation are performed automatically as explained below. For the examples shown here the mean pixel size in object space was ~0.8 mm.

■ A green line laser with adjustable focus, allowing width of the laser line ~0.5 mm.

■ A calibration board with a typical black-and-white chessboard pattern, with one of its squares being red.

■ Two planes forming a corner (optional).

2. STEPS OF THE SCANNING PROCESS

2.1 System calibration

In the calibration process the parameters of the interior and the (scaled) relative orientation of the two web cameras are estimated automatically, by processing the synchronized images of a chess-board. As documented in detail in Douskos et al. (2008), our group has developed a toolbox for automatic camera calibration*. It uses images of a typical planar chess-board pattern (square black-and-white tiles), on which feature points are extracted with Harris operator. Among extracted points, all pattern nodes are identified and then ordered in rows and columns, thus allowing camera calibration via bundle adjustment (no physical point correspondences are needed since exterior orientation is in this case irrelevant).

For the purposes of this study the algorithm had to be modified in several respects, besides the requirement for pattern squares of known size for scaling. Unknown are the interior orientations of two cameras, the pose of one camera and the (scaled) relative orientation of the camera pair. Also, a pattern coordinate system should be fixed to ensure true image point correspondences. To this end, one pattern square is red and needs to be visible in all images. For each image the green channel is subtracted from the red one. After thresholding and dilation, edges are extracted. In contrast to possible noisy polygons, the polygon around the red square is that which contains four extracted Harris points. These points allow identifying the origin and orientation of the coordinate axes of the calibration pattern. They also give good estimates for the grid step in the vicinity of the red square. The nodes are then identified and ordered in correspondence to the pattern nodes, which allows camera calibration with bundle adjustment (initial values are found after Douskos et al., 2008). Fig. 2 gives a view of the calibration pattern and the extracted nodes.

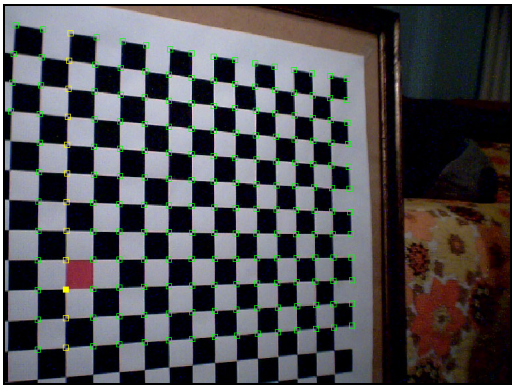

Figure 3. Image of the calibration pattern, with identified origin, points of the pattern y-axis (yellow) and all other nodes (green).

In all calibration tests the standard error of the adjustment was below 0.2 pixels. Table 3 shows the calibration data for our ex-

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

perimental application (k and p are coefficients of the radial and decentering lens distortions, respectively); image skewness has been ignored in this case. The cameras were mounted with axes considerably convergent (rotation angle φ).

Table 3
Calibration results (from 18 image pairs)

$\sigma_o = 0.19$ pixel		
	left camera	right camera
c_x (pix)	944.59 ± 0.22	959.43 ± 0.20
c_y (pix)	942.19 ± 0.22	955.82 ± 0.22
x_o (pix)	–19.12 ± 0.46	–65.28 ± 0.48
y_o (pix)	–2.39 ± 0.31	–2.61 ± 0.33
$k_1(\times 10^{-07})$	–1.40 ± 0.02	–1.60 ± 0.02
$k_2(\times 10^{-14})$	–5.34 ± 1.62	–5.24 ± 0.95
$p_1(\times 10^{-06})$	2.09 ± 0.19	–3.90 ± 0.19
$p_2(\times 10^{-06})$	–2.61 ± 0.13	–0.01 ± 0.13
	relative orientation	
B_x (cm)	38.10 ± 0.01	
B_y (cm)	2.17 ± 0.00	
B_z (cm)	–11.55 ± 0.02	
ω (°)	–6.13 ± 0.03	
φ (°)	33.72 ± 0.04	
κ (°)	13.13 ± 0.01	

2.2 Image acquisition and image subtraction

For scanning, the images are taken simultaneously with the web cameras. Illumination conditions have to be such that the object texture is captured, yet not too bright to allow detecting change between object surface and laser profile. Dull surfaces produce better scans; for more shiny surfaces, scanning with no exterior light source is recommended. To isolate a laser profile from the image background, a reference image is usually subtracted from the actual image. A reference image may be acquired separately (Winkelbach et al., 2006) or generated with a temporal median approach using a small number of images, as it is the case here. A further use of these background images is to finally provide each 3D point with its specific photo-texture.

2.3 Epipolar transformation

With the calibration and orientation data, image pairs acquired during scanning may be normalized, i.e. resampled to epipolar images (the resulting images are error-free since the known lens distortion effects can be removed); thus, the search for point homologies is confined on corresponding epipolar lines. In Fig. 4 an image pair and its epipolar images are presented.

2.4 Peak detection

The points to be found on epipolar lines are their intersections with the viewed laser profiles, i.e. it is needed to detect the peak (or peaks) in the direction of each image row. Of course, the accuracy in 3D reconstruction with such scanners greatly depends upon the accuracy with which a laser stripe is detected. Several peak detection approaches have been reported (Fisher & Naidu, 1996). In general, it is considered that recorded laser stripes can be adequately approximated with Gaussian profiles. Thus, Blais & Rioux (1986) and Forest et al. (2004) detect the position of a peak (maximum light intensity) at the zero-crossing of the first derivative of the Gaussian curve. In this context, these authors propose filtering techniques for noise removal, also taking into account different optical properties of the surfaces.

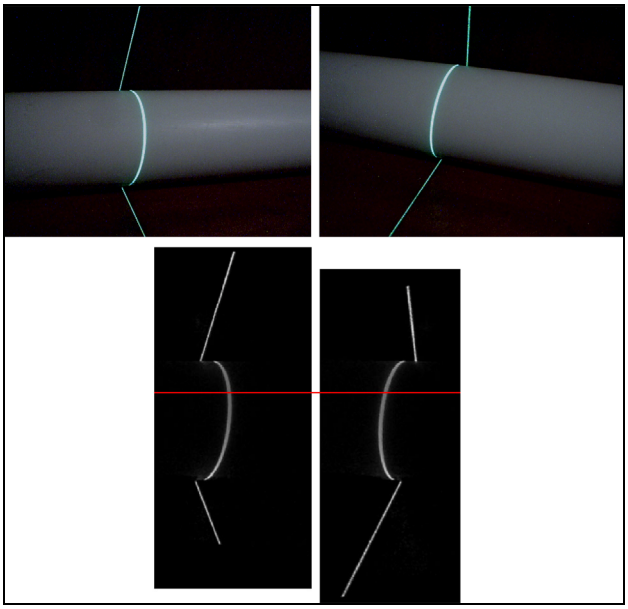

Figure 4. Image pair (above) and epipolar subtracted images.

Here, a Gaussian curve was directly fitted to the profile. It was found that, for the surfaces scanned in this study, a simple preprocessing step suffices for noise removal. For the current setup (good quality laser) this is done with a 3×3 Gaussian filter. For lasers of poorer quality it should be preceded by a median filter.

First, a threshold is applied on the image row to give estimates for the position of the peak and the width of the occurrence. On each row there might be more than one occurrence (see below). All peaks of a row are then interpolated, provided that the stripe width is below a threshold. This excludes 'stretched' stripes due to laser planes being close to tangent to a curved object surface. The position of a peak in the x-direction is at the maximum of the Gaussian curve fitted to the gray value data:

$$f(x) = ae^{-\frac{(x-b)^2}{2\sigma^2}} + d \qquad \text{(Eq. 1)}$$

In the equation, parameter *b* provides the estimated subpixel position of the peak. It is advisable, however, to relax the strictness of this 1D interpolation and exploit additional information from the neighbourhood of the initially approximated location. Thus, three Gaussian profiles are simultaneously fitted in an adjustment with a common *b* parameter in the image x-direction. The additional curves (participating with smaller weight) are interpolated in the two diagonal directions through the initial peak approximation. Result is peak estimates still on epipolar lines, but with x-coordinates influenced by gray value distribution in more directions (the curve in Fig. 5 is slightly shifted along the epipolar line under the influence of the diagonal directions). It was found that this process helps reduce the effect of noise, providing peaks which lead to more accurate 3D reconstruction.

2.5 Triangulation

First, we will refer to the rows which produced only one pair of x-coordinates (one peak in each image). Hence, available for all these extracted point correspondences are one y-coordinate, that of the epipolar line, and a coordinate pair (x, x′).

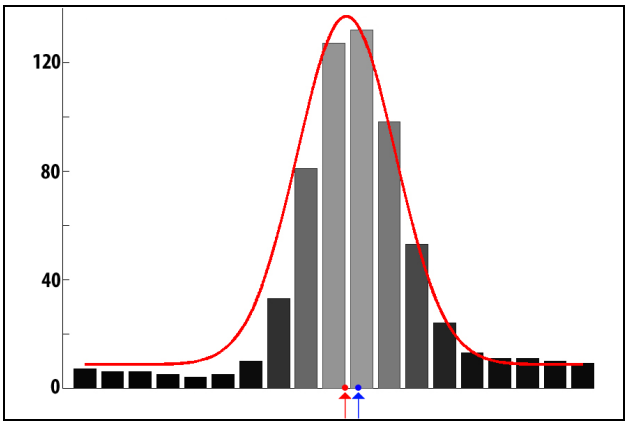

Figure 5. Curve fitting to a gray value profile along an epipolar line with diagonal directions taken into account (peak location when the fit refers only to the epipolar line is seen in blue).

2.5.1 Background planes. Prior to object scanning, the corner formed by the two background planes is scanned (Fig. 6). From image coordinates x, x′, y, ground points XYZ are reconstructed with the simple parallax equations, under a constraint that they are coplanar (on the laser plane). Using RANSAC, two 3D lines are then fitted to these 3D points; thus, for each image pair the points of the two planes are identified. To the two point clouds, which result from all pairs of frames, planes are finally fitted to provide equations of the two background planes.

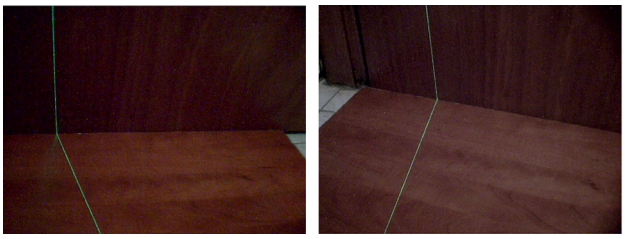

Figure 6. Image pair of the background planes.

2.5.2 Reconstruction algorithm. The initial step of calculating 3D coordinates for all points by means of the parallax equations is repeated for each pair of frames in the scanning phase. From these 3D points it is possible to estimate laser plane coefficients, and also to separate the points in three groups. Besides the two groups with the points of the background planes, the third group includes only those which pertain to the scanned object surface. This separation is done using for each point a distance threshold to the background planes.

The reconstruction algorithm relies on typical photogrammetric triangulation (parallax equations), on which two geometric constraints are imposed to strengthen the reliability of the solution:

- All points reconstructed from the laser stripe of each image pair should belong to the laser plane, for which the optimal coefficients are calculated.
- All points ascribed to the two background planes should, additionally, satisfy the corresponding plane equation. It is not advisable to handle this constraint as rigid – the fitted planes are subject to the accuracy limits of the approach – but rather as a 'soft' constraint with large weight.

It would be feasible to perform a unified solution with the background planes as common elements. In our current implementation it is all points of a stripe which are calculated in one step through a least squares adjustment. The total of these 3D points constitutes the point cloud as generated from a specific camera position. An extra step is to back-project all 3D points onto the two reference images and interpolate RGB sets; their average complements the 3D data to produce a XYZ–RGB set (Fig. 7).

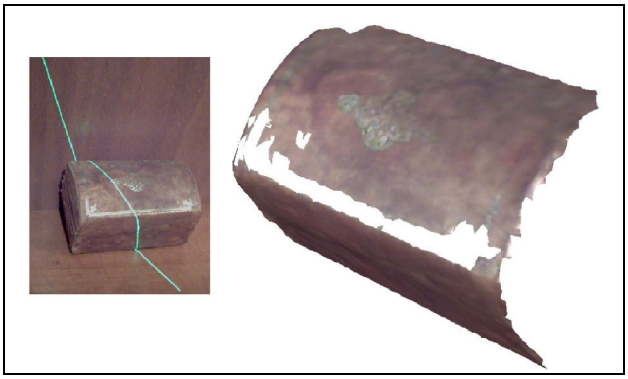

Figure 7. Photo-textured model from a single scan.

In the actual scanning process it is advisable to scan the surface more than once from each view. This allows generating denser point clouds, if necessary, but it would also provide the data for possibly filtering out blunders in depth estimation. The user has to repeat the scanning process from other camera views in order to fully capture the desired parts (or the whole) of a surface. In a final step, individual scans must be processed and registered to produce the final 3D model. In our case this was done using available modeling software (for a discussion on the question of registration in this context see e.g. Winkelbach et al., 2006).

2.5.3. Multiple peaks. As mentioned, in the peak detection step epipolar lines with more than one peak are encountered. This is due to surface relief (occlusions) or e.g. to reflected laser lines. These rows are stored. After the adjustment for a laser stripe, all possible combinations of peaks on such epipolar lines give 3D points, some of which are actual object points; accepted as valid surface points are those which meet a certain distance threshold from the estimated laser plane. Example of a double peak due to occlusions is shown in Fig. 8.

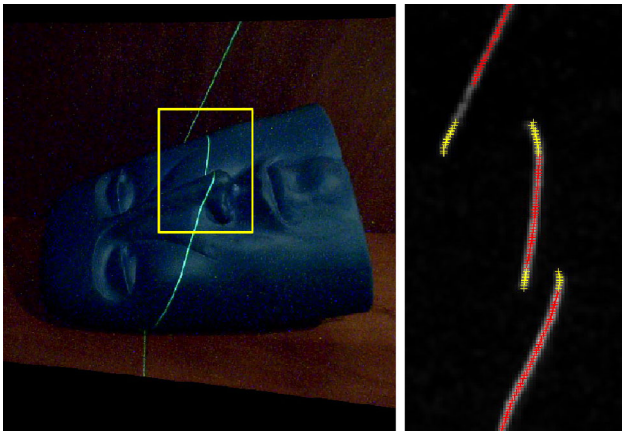

Figure 8. Single (in red) and double (yellow) peak occurrences due to occlusions.

3. EXPERIMENTAL EVALUATION

3.1 Expected accuracy

Epipolar geometry refers to the 'normal case' of the stereo pair; thus, the accuracy of the method may be estimated theoretically

in the simple terms of the parallax equation. The error σ_p of the x-parallax $p = x - x'$ is propagated to estimated depth through the image scale and the base-to-distance ratio. The σ_p values are the result of the uncertainty σ_x in the x-direction of peak positions estimated through Gaussian curve fitting (parameter *b* in Eq. 1). Here, σ_x values were mostly below 0.1 pixels. For instance, for a typical laser stripe with 400 established point homologies, the mean σ_x value for the image pair was 0.086 and 0.081 pixels. In the range 0.2–0.6 pixels were 2.8% of the σ_x values; 10.2% of them were between 0.1–0.2 pixels; 87% were below 0.1 pixels.

Therefore, a value $\sigma_x = 0.1$ pixels might be assumed, and hence $\sigma_p = 0.15$ pixels. The imaging distance in our experiments was 65–90 cm. Using c = 950 pixels and B = 40 cm (see Table 2), a realistic expectation for the depth estimation accuracy, referring to these imaging distances, would be $\sigma_z = 0.15$–0.30 mm.

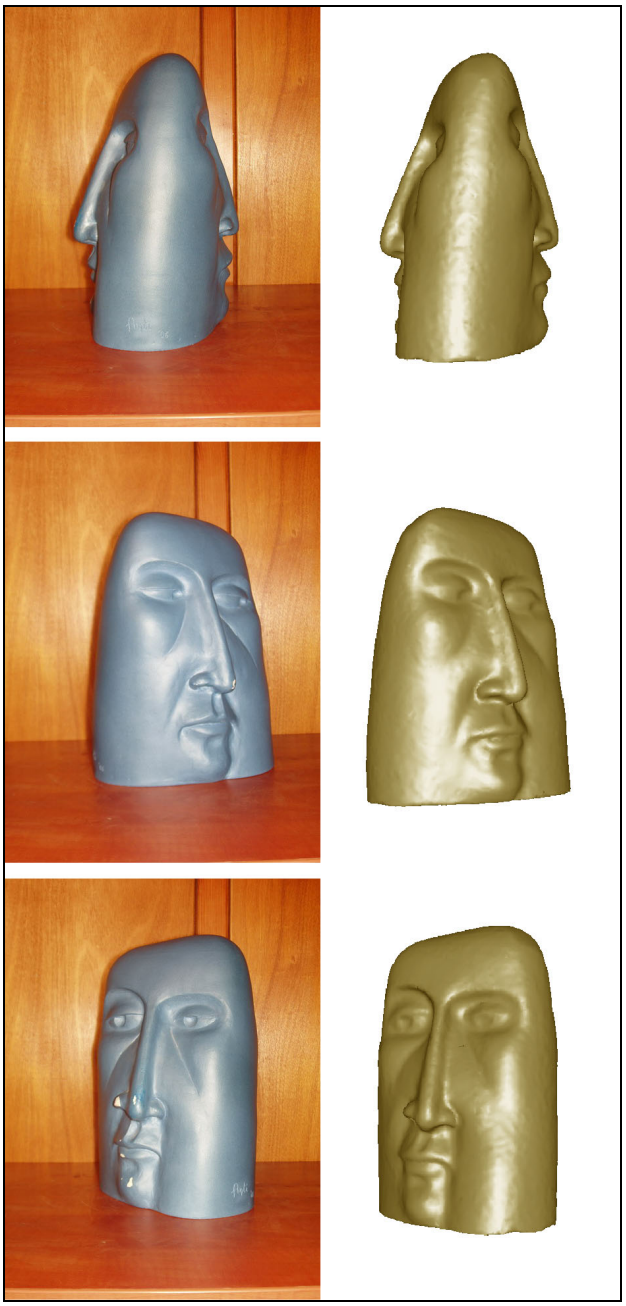

Figure 9. Different views of the object (left) and corresponding views of the rendered 3D model (right).

3.2 Evaluation of accuracy

The experimental results were essentially in agreement with this estimation. Plane-fitting adjustments in all tests performed here have yielded standard errors 0.14–0.26 mm. Next, a white PVC pluming tube with a nominal diameter of 125 mm was scanned (Fig. 4). Using its known diameter, a cylinder was then fitted to the 26,200 XYZ values of the point cloud, extracted from 130 pairs of frames. This represented approximately 2/5 of the perimeter, being the product from a single scanning position. The standard error of the surface-fitting adjustment – from which only very few obvious outliers were excluded – was 0.17 mm.

3.3 Practical application

The precision of the method was also assessed in the case of a practical application. A sculpture (35 cm in height) was scanned from 11 different positions. More than one scans were made at each position to check the scanning repeatability. The average distance between scans was 0.17–0.28 mm. Finally, the separate scans were merged in a single 3D model of the surface, which is presented in Fig. 9.

4. CONCLUSION

A low-cost method has been reported for capturing 3D scan data from different viewpoints, which can then be merged in a single 3D model of the scanned surface. Stereovision is combined with the slit scanner principle, to allow introducing extra coplanarity constraints. Given the precise calibration process, this may have improved accuracy itself only slightly, but has added in robustness. The accuracy in 3D reconstruction has been checked both theoretically and experimentally, and was found above 0.3 mm with the particular setup. Main future tasks include the development of further means for automatically detecting outliers, both in the peak detection and the reconstruction steps, and the experimentation with surfaces of different optical properties.

Acknowledgement. *The authors wish to thank sculptor Antigoni Pantazi (www.antipantazi.com) for releasing the 'Head' of Fig. 9 (ceramic sculpture, stoneware, terra sigillatta).*

REFERENCES

Blais F., Rioux M., 1986. Real-time numerical peak detector. Signal Processing, 11(2), pp. 145-155.

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.

D'Apuzzo N., 1998. Automated photogrammetric measurement of human faces. Int. Arch. Photogram. Rem. Sens., 32 (B5), pp. 402-407.

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.

Fisher R.B., Ashbrook A.P., Robertson C., Werghi N., 1999. A low-cost range finder using a visually located, structured light source. Proc. 2nd Int. Conf. on 3D Digital Imaging and Modeling (3DIM '99), pp. 24-33.

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.

Forest J., Salvi J., Cabruja E., Pous C., 2004. Laser stripe peak detector for 3D scanners. A FIR filter approach. Proc. 17th Int. Conf. on Pattern Recognition (ICPR'04), vol. 3, pp. 646-649.

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.

Salvi J., Pagès J., Batlle J., 2004. Pattern codification strategies in structured light systems. Pattern Recognition, 37(4), pp. 827–849.

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.