3D PHOTOGRAPHY USING SHADOWS

Jean-Yves Bouguet† and Pietro Perona†‡
† California Institute of Technology, 136-93, Pasadena, CA 91125, USA
‡ Università di Padova, Italy
{bouguetj,perona}@vision.caltech.edu

ABSTRACT

A simple and inexpensive approach for extracting the three-dimensional shape of objects is presented. It is based on 'weak structured lighting;' and requires little hardware besides the camera: a light source (a desk-lamp or the sun), a stick and a checkerboard. The object, illuminated by the light source, is placed on a stage composed of a ground plane and a back plane; the camera faces the object. The user moves the stick in front of the light source, casting a moving shadow on the scene. The 3D shape of the object is extracted from the spatial and temporal location of the observed shadow. Experimental results are presented on three different scenes (indoor with a desk lamp and outdoor with the sun) demonstrating that the error in reconstructing the surface is less than 0.5% of the size of the object.

1 INTRODUCTION AND MOTIVATION

One of the most valuable functions of our visual system is informing us about the shape of the objects that surround us. Ever-faster computers, progress in computer graphics, and the widespread expansion of the Internet have recently generated much interest in imaging both the geometry and surface texture of objects. The applications are numerous (animation and entertainment, industrial design, archiving, virtual visits to museums...).

In designing a system for recovering shape, different engineering tradeoffs are proposed by each application. The main parameters to be considered are: cost, accuracy, ease of use and speed of acquisition. So far, the commercial 3D scanners (e.g. the Cyberware scanner) have emphasized accuracy over the other parameters. These systems use motorized transport of the object, and active (laser, LCD projector) lighting of the scene, which makes them very accurate, but unfortunately expensive and bulky [1, 8, 9].

An interesting challenge is to take the opposite point of view: emphasize low cost and simplicity and design 3D scanners that demand little more hardware than a PC and a video camera by making better use of the data that is available in the images.

We propose a method for capturing 3D surfaces that is based on what we call 'weak structured lighting.' It yields good accuracy and requires minimal equipment besides a computer and a camera: a stick, a checkerboard, and a point light source. The light source may be a desk lamp for indoor scenes. A human operator, acting as a low precision motor, is also required.

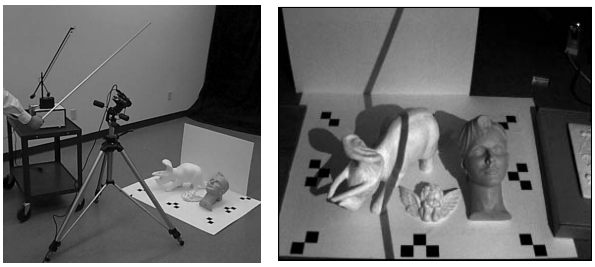

Figure 1: **The general setup of the proposed method:** The camera is facing the scene illuminated by the light source (left). The figure illustrates an indoor scenario where a desk lamp (without reflector) is used as light source. Outdoors the lamp is substituted by the sun. The objects to be scanned are positioned on the ground floor (horizontal plane), in front of a background plane. When an operator freely moves a stick in front of the light, a shadow is cast on the scene. The camera acquires a sequence of images $I(x, y, t)$ as the operator moves the stick so that the shadow scans the entire scene. A sample image is shown on the right figure. This constitutes the input data to the 3D reconstruction system.

We start with a description of the method in Sec. 2, followed in Sec. 3 by a number of experiments that assess the convenience and accuracy of the system in indoor as well as outdoor scenarios. We end with a discussion and conclusions in Sec. 4.

2 DESCRIPTION OF THE METHOD

The general principle consists of casting a moving shadow with a stick onto the scene, and estimating the three dimensional shape of the scene from the sequence of images of the deformed shadow. Figure 1 shows a typical setup of the method. The objective is to extract scene depth at every pixel in the image. The point light source and the stick define, at every time instant, a plane; therefore, the boundary of the shadow that is cast by the stick on the scene is the intersection of this plane with the surface of the object. We exploit this geometrical insight for reconstructing the 3D shape of the object. Figure 2 gives the geometrical principle of the method. Notice that if the light source is at a known location in space, then the shadow plane $\Pi(t)$ may be directly inferred from the point S and the line $\Lambda_h(t)$. Consequently, in such cases, the additional plane $\Pi_v(t)$ is not required.

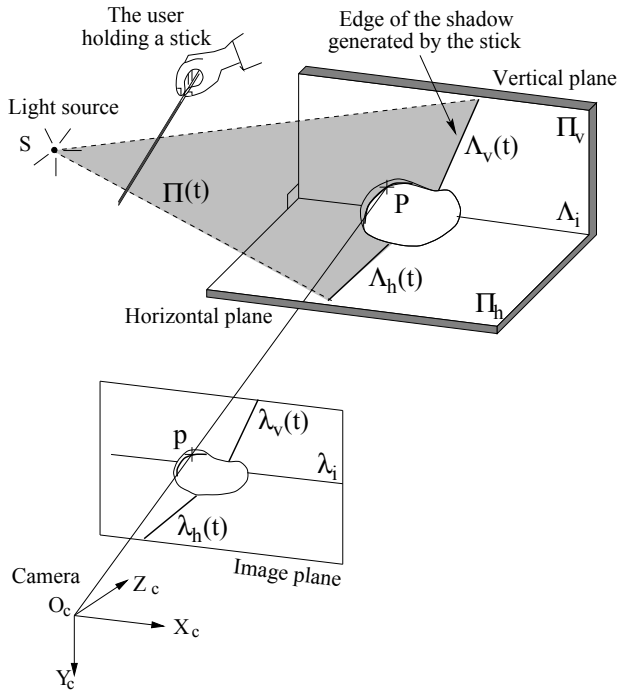

Figure 2: **Geometrical principle of the method:** Approximate the light source with a point S, and denote by Π_h the horizontal plane (ground) and Π_v a vertical plane orthogonal to Π_h. Assume that the position of the plane Π_h in the camera reference frame is known from calibration. We infer the location of Π_v from the projection λ_i (visible in the image) of the intersection line Λ_i between Π_h and Π_v. The goal is to estimate the 3D location of the point P in space corresponding to every pixel p (of coordinates $\overline{x}_c$) in the image. Call t the time when the shadow boundary passes by a given pixel $\overline{x}_c$ (later referred to as the *shadow time*). Denote by $\Pi(t)$ the corresponding shadow plane at that time t. Assume that two portions of the shadow projected on the two planes Π_h and Π_v are visible on the image: $\lambda_h(t)$ and $\lambda_v(t)$. After extracting these two lines, we deduce the location in space of the two corresponding lines $\Lambda_h(t)$ and $\Lambda_v(t)$ by intersecting the planes $(O_c, \lambda_h(t))$ and $(O_c, \lambda_v(t))$ with Π_h and Π_v respectively. The shadow plane $\Pi(t)$ is then the plane defined by the two non-collinear lines $\Lambda_h(t)$ and $\Lambda_v(t)$. Finally, the point P corresponding to $\overline{x}_c$ is retrieved by intersecting $\Pi(t)$ with the optical ray (O_c, p). This final stage is called triangulation. Notice that the key steps are: (a) estimate the shadow time $t_s(\overline{x}_c)$ at every pixel $\overline{x}_c$ (*temporal processing*), (b) locate the two reference lines $\lambda_h(t)$ and $\lambda_v(t)$ at every time instant t (*spatial processing*), (c) determine the shadow plane, and (d) triangulate and calculate depth.

2.1 Calibration

The goal of calibration is to recover the location of the two planes Π_h and Π_v and the *intrinsic* camera parameters (focal length, optical center and radial distortion factor). The procedure consists of first placing a planar checkerboard pattern on the ground in the location of the objects to scan (see figure 3-left). From the image captured by the camera (figure 3-right), we infer the intrinsic and extrinsic parameters of the camera, by matching the projections onto the image plane of the known grid corners with the expected projection directly measured on the image (extracted corners of the grid); a method very much inspired by the algorithm proposed by Tsai in [10]. A description of the whole procedure can be found in [3].

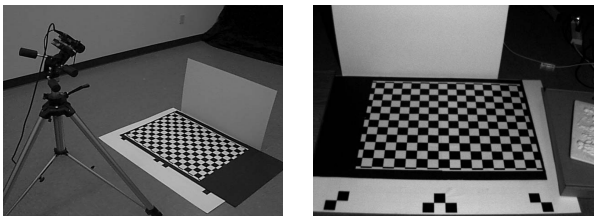

Figure 3: **Camera calibration**

2.2 Spatial and temporal shadow edge localization

A fundamental stage of the method is the detection of the lines of intersection of the shadow plane $\Pi(t)$ with the two planes Π_h and Π_v; a simple approach to extract $\overline{\lambda}_h(t)$ and $\overline{\lambda}_v(t)$ may be used if we make sure that a number of rows at the top and bottom of the image are free from objects. Then the two tasks to accomplish are: **(a)** Localize the edges of the shadow that are directly projected on the two orthogonal planes $\lambda_h(t)$ and $\lambda_v(t)$ at every time instant t (every frame), leading to the set of all shadow planes $\Pi(t)$, **(b)** Estimate the time $t_s(\overline{x}_c)$ (*shadow time*) where the edge of the shadow passes through any given pixel $\overline{x}_c = (x_c, y_c)$ in the image. Curless and Levoy [6] demonstrated that such a spatio-temporal approach is appropriate to preserve sharp discontinuities in the scene. A similar temporal processing for range sensing was used by Gruss, Tada and Kanade in [8]. Details of our implementation are given in figure 4 and in references [3, 4]. Notice that the right edge of the shadow corresponds to the front edge of the temporal profile, because the shadow was scanned from left to right in all experiments. Intuitively, pixels corresponding to occluded regions in the scene do not provide any relevant depth information. Therefore, we only process pixels with contrast value $I_{\text{contrast}}(x, y) \doteq I_{\max}(x, y) - I_{\min}(x, y)$ larger than a predefined threshold I_{thresh} (set to 30 in all experiments reported in this paper).

Although the two time-global parameters $I_{\min}$ and $I_{\max}$ are needed to compute I_{shadow}, an implementation of the algorithm exists that does not require storage of the complete image sequence in memory and therefore allows for real-time implementations (see [4]).

2.3 Triangulation

Once the shadow time $t_s(\overline{x}_c)$ is estimated at a given pixel $\overline{x}_c$, one can identify the corresponding shadow plane $\Pi(t_s(\overline{x}_c))$. Then, the 3D point P associated to $\overline{x}_c$ is retrieved by intersecting $\Pi(t_s(\overline{x}_c))$ with the optical ray $(O_c, \overline{x}_c)$ (see figure 2). Notice that the shadow time $t_s(\overline{x}_c)$ acts as an index to the shadow plane list $\Pi(t)$. Since $t_s(\overline{x}_c)$ is estimated at sub-frame accuracy, the final plane $\Pi(t_s(\overline{x}_c))$ actually results from linear interpolation between the two planes $\Pi(t_0 - 1)$ and $\Pi(t_0)$ if $t_0 - 1 < t_s(\overline{x}_c) < t_0$ and t_0 integer. Once the range data are recovered, a mesh may be generated by connecting neighboring points in triangles. Rendered views of three reconstructed surface structures can be seen in figures 5, 6 and 7.

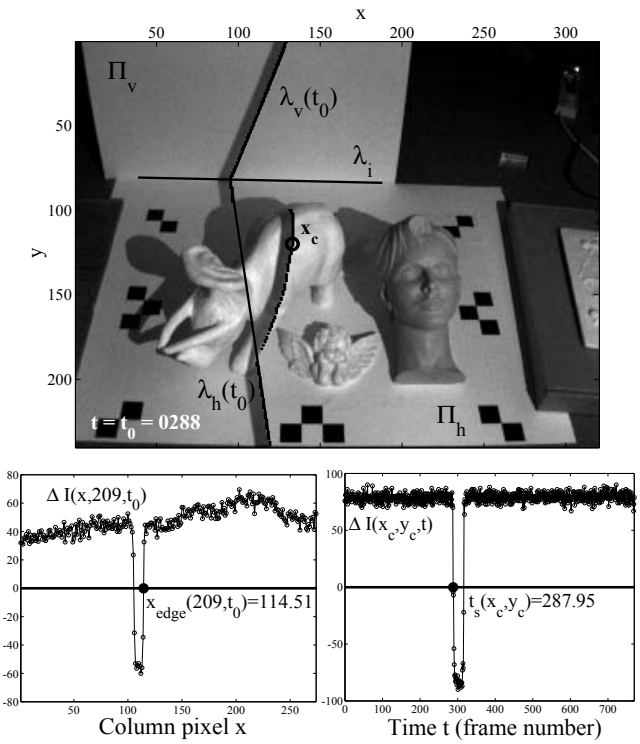

Figure 4: **Spatial and temporal shadow localization**: The first step consists of localizing spatially the shadow edges $\lambda_h(t)$ and $\lambda_v(t)$ at every integer time t_0 (i.e. every frame) using a number of rows on the image free from occluding objects. The second processing step consists of extracting at every pixel $\overline{x}_c$, the time $t_s(\overline{x}_c)$ of passage of the shadow edge (top figures). For any given pixel $\overline{x}_c = (x, y)$, define $I_{\min}(x, y) \doteq \min_t (I(x, y, t))$ and $I_{\max}(x, y) \doteq \max_t (I(x, y, t))$ as its minimum and maximum brightness throughout the entire sequence. We then define the shadow edge to be the locations (in space-time) where the image $I(x, y, t)$ intersects with the threshold image $I_{\text{shadow}}(x, y) \doteq (I_{\min}(x, y) + I_{\max}(x, y))/2$. This may be also regarded as the zero crossings of the difference image $\Delta I(x, y, t) \doteq I(x, y, t) - I_{\text{shadow}}(x, y)$. The two bottom plots illustrate the shadow edge detection in the spatial domain (to find $\lambda_h(t)$ and $\lambda_v(t)$) and in the temporal domain (to find $t_s(\overline{x}_c)$). The bottom-left figure shows the profile of $\Delta I(x, y, t)$ along row $y = 209$ at time $t = t_0 = 288$ versus the column pixel coordinate x. The second zero crossing of that profile corresponds to one point $\overline{x}_{\text{edge}}(t_0) = (114.51, 209)$ belonging to $\lambda_h(t_0)$ (computed at subpixel accuracy by linear interpolation). Identical processing is applied on 39 other rows for $\lambda_h(t_0)$ and 70 rows for $\lambda_v(t_0)$ in order to retrieve the two edges (by least squares line fitting across the two sets of points on the image). Similarly, the bottom-right figure shows the temporal profile $\Delta I(x_c, y_c, t)$ at the pixel $\overline{x}_c = (x_c, y_c) = (133, 120)$ versus time t (or frame number). The shadow time at that pixel is defined as the first zero crossing location of that profile: $t_s(133, 120) = 287.95$ (computed at sub-frame accuracy by linear interpolation).

3 EXPERIMENTAL RESULTS

3.1 Calibration accuracy

For a given setup, we acquired 5 images of the checkerboard pattern (see figure 3-right), and performed independent calibrations on them. The checkerboard, placed at different positions in each image, consisted of 187 visible corners on a 16×10 grid. We computed both mean values and standard deviations of all the parameters independently: the focal length f_c, radial distortion factor k_c and ground plane position Π_h. Regarding the ground plane position, it is convenient to look at its distance d_h to the camera origin O_c and its normal vector $\overline{n}_h$ expressed in the camera reference frame. The following table summarizes the calibration results:

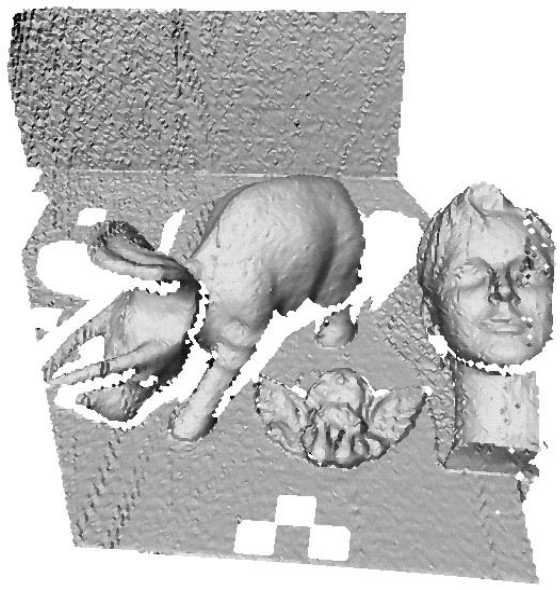

Figure 5: **Experiment 1 - Indoor scene**

Parameters	Estimates	Relative errors
f_c (pixels)	853.7 ± 1.5	0.2%
k_c	-0.233 ± 0.002	1%
d_h (cm)	112.1 ± 0.1	0.1%
$\overline{n}_h$	$\begin{pmatrix} -0.0529 \pm 0.0003 \\ 0.7322 \pm 0.0003 \\ 0.6790 \pm 0.0003 \end{pmatrix}$	0.05%

This accuracy is sufficient for not inducing any significant global distortion onto the final recovered shape (see [4] for discussions).

3.2 Scene reconstructions

In the first indoor reconstructed scene (figure 5), the surface noise was estimated to approximately 0.5 mm in standard deviation over 50 cm large objects (a relative reconstruction error of 0.1%). Figure 6 shows the result of an outdoor scanning using with the sun as light source. The surface error in that scene is approximately 1 mm, or equivalently a relative error of approximately 0.5%. Figure 7 shows the reconstruction results on scanning a car with the sun. The reconstruction errors were estimated to approximately 1cm, leading to 0.5% of the size of the car. The larger errors in the two last experiments may be explained by the fact that the sun is not an ideal point light source (this is subject to further investigations). Notice that there were not significant global deformation in both reconstructions which leads us to believe that calibration provides sufficiently accurate estimate of the geometry (see [4]).

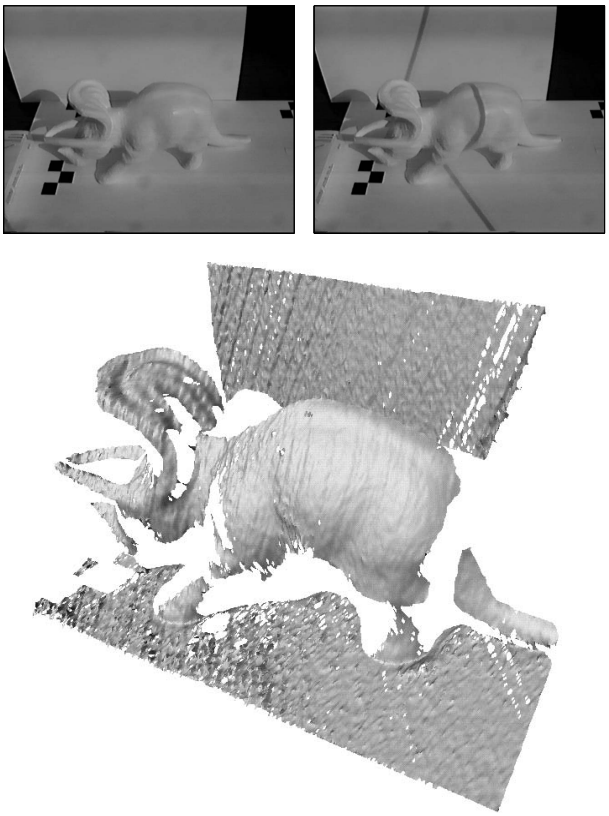

Figure 6: **Experiment 2 - Outdoor scanning of an object**

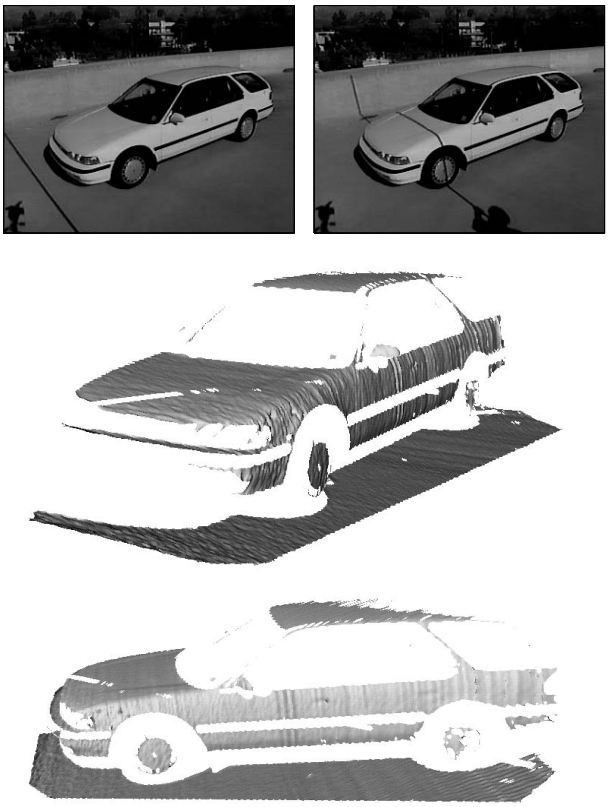

Figure 7: **Experiment 3 - Outdoor scanning of a car**

4 CONCLUSION AND FUTURE WORK

We have presented a simple, low cost system for extracting surface shape of objects. In case of outdoor scenarios, the sun may be used as light source that is allowed to move during a scan. The method requires very little processing and image storage so that it can be implemented in real time. The accuracies that we obtained on the final reconstructions are reasonable (at most 0.5% of the size of the scene). In addition, the final outcome is a dense and organized coverage of the surface (one point in space for each pixel in the image), allowing direct texture mapping. This system may be used as a front end acquisition technique for complete 3D object modeling. One may take multiple scans of the object at different locations in space, and then align the sets of range images [2, 5, 7, 11].

Acknowledgments

This work is supported in part by the California Institute of Technology; an NSF National Young Investigator Award to P.P.; a STC fund; the Center for Neuromorphic Systems Engineering funded by the National Science Foundation at the California Institute of Technology; and by the California Trade and Commerce Agency, Office of Strategic Technology. We wish to thank all the colleagues that helped us throughout this work, especially Peter Schröder, Paul Debevec, and Luis Goncalves for very useful discussions.

References

[1] Paul Besl, *Advances in Machine Vision*, chapter 1 - Active optical range imaging sensors, pages 1–63, Springer-Verlag, 1989.

[2] P.J. Besl and N.D. McKay, "A method for registration of 3-d shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[3] Jean-Yves Bouguet and Pietro Perona, "3D photography on your desk", Technical report, California Institute of Technology, 1997, available at: http://www.vision.caltech.edu/bouguetj/ICCV98.

[4] Jean-Yves Bouguet and Pietro Perona, "3D photography on your desk", *Proc. 6^{th} Int. Conf. Computer Vision*, pages 43–50, 1998.

[5] C.L. Bajaj, F. Bernardini, and G. Xu, "Automatic reconstruction of surfaces and scalar fields from 3D scans", *In SIGGRAPH '95, Los Angeles, CA*, pages 109–118, August 1995.

[6] Brian Curless and Marc Levoy, "Better optical triangulation through spacetime analysis", *Proc. 5^{th} Int. Conf. Computer Vision*, pages 987–993, 1995.

[7] Brian Curless and Marc Levoy, "A volumetric method for building complex models from range images", *SIGGRAPH96, Computer Graphics Proceedings*, 1996.

[8] A. Gruss, S. Tada, and T. Kanade, "A VLSI Smart Sensor for Fast Range Imaging", In *DARPA93*, pages 977–986, 1993.

[9] Marjan Trobina, "Error model of a coded-light range sensor", Technical Report BIWI-TR-164, ETH-Zentrum, 1995.

[10] R. Y. Tsai, "A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses", *IEEE J. Robotics Automat.*, RA-3(4):323–344, 1987.

[11] G. Turk and M. Levoy, "Zippered polygon meshes from range images", *In SIGGRAPH '94*, pages 311–318, July 1994.