

3D PHOTOGRAPHY ON YOUR DESK

Students: Jean-Yves Bouguet
Faculty: Pietro Perona
Support: NSF, NYI, State of California

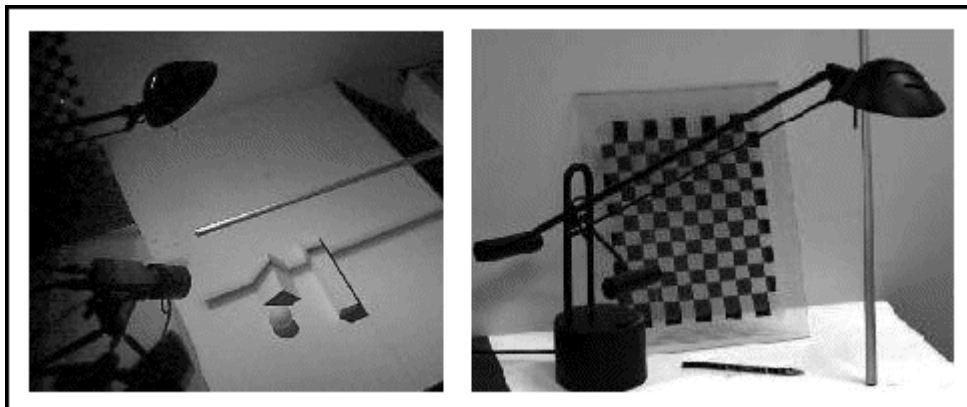
MOTIVATION & AIMS:

One of the most valuable functions of our visual system is informing us about the shape of the objects that surround us. Manipulation, recognition, and navigation are amongst the tasks that we can better accomplish by seeing shape. Ever-faster computers, progress in computer graphics, and the widespread expansion of the Internet have recently generated much interest in systems that may be used for imaging both the geometry and surface texture of object. The applications are numerous. Perhaps the most important ones are animation and entertainment, industrial design, archiving, virtual visits to museums and commercial on-line catalogues.

We propose a method for capturing 3D surfaces that is based on 'weak structured lighting' [1]. It yields good accuracy and requires minimal equipment besides a computer and a camera: a pencil, a checkerboard and a desk-lamp -- all readily available in most homes; some intervention by a human operator, acting as a low precision motor, is also required.

RESEARCH:

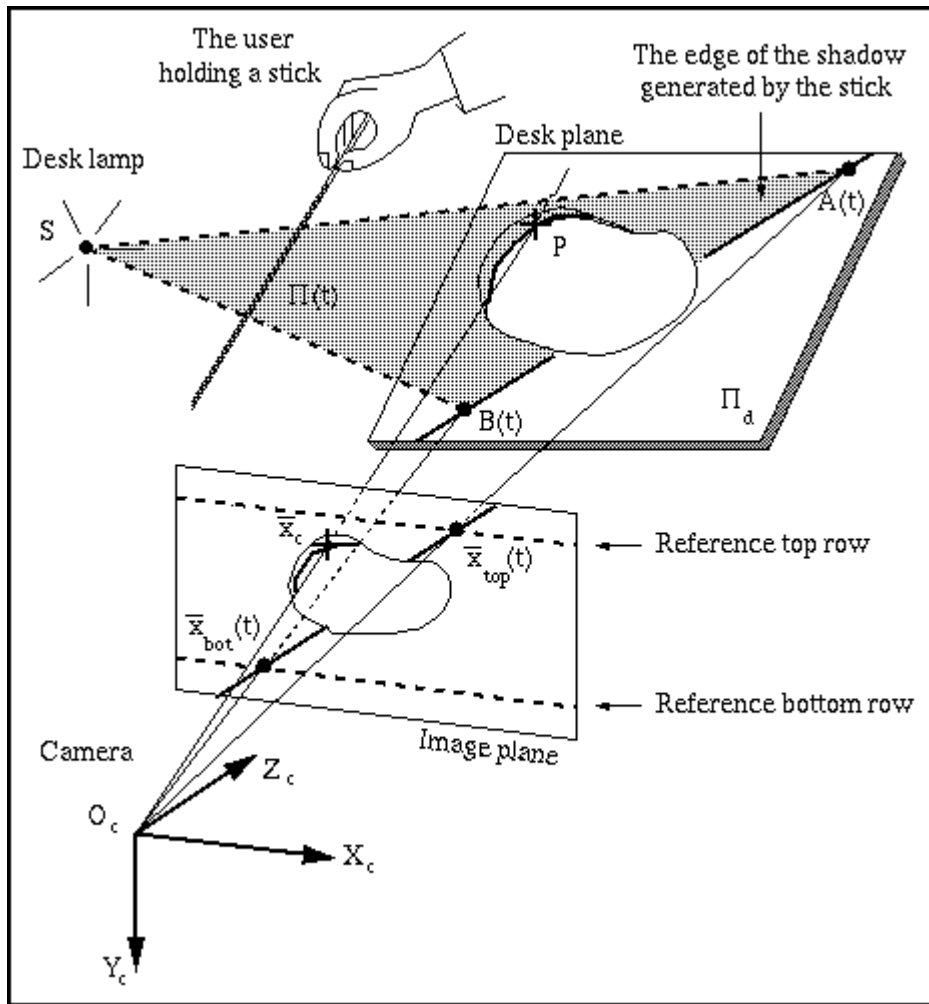
THE IDEA:



The camera is facing the scene illuminated by a halogen desk lamp (left). The scene consists of objects on a plane (the desk). When an operator freely moves a stick in front of the lamp (over the desk), a shadow is cast on the scene. The camera acquires a sequence of images as the operator moves the stick so that the shadow scans the entire scene. This constitutes the input data to the 3D reconstruction system. The three dimensional shape of the scene is reconstructed using the spatial and temporal properties of the shadow boundary throughout the input sequence. The right-hand figure shows the necessary equipment besides the camera: a desk lamp, a calibration grid and a pencil for calibration, and a stick for the shadow. One could use the pencil instead of the stick.

GEOMETRICAL DESCRIPTION:

The general principle consists of casting a shadow onto the scene with a pencil or another stick, and using the image of the deformed shadow to estimate the three dimensional shape of the scene. The objective is to extract scene depth at every pixel in the image.



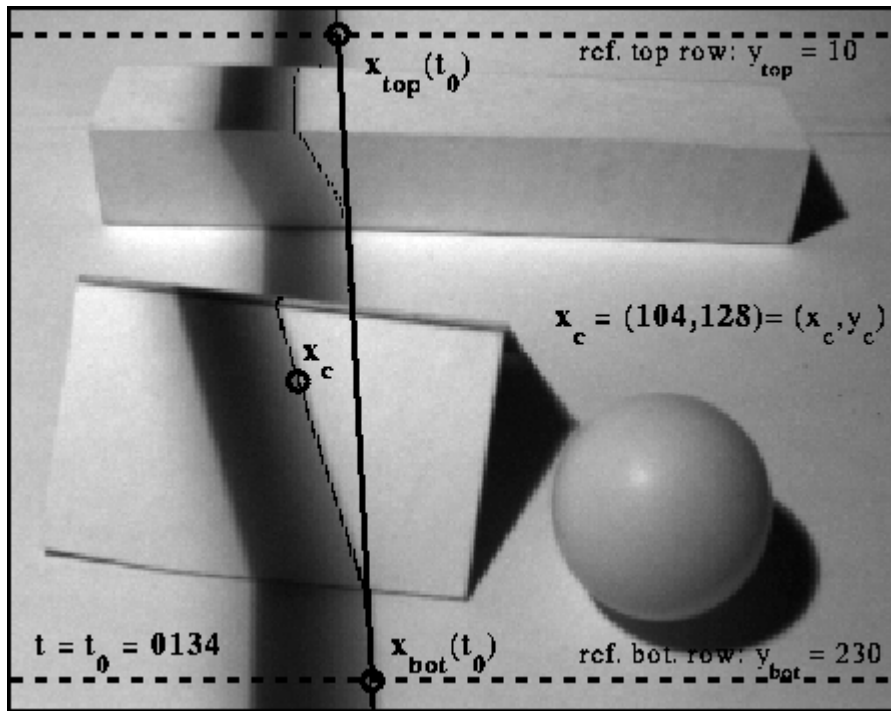
The goal is to estimate the 3D location of the point P in space corresponding to every pixel x_c in the image. Call t the time at which a given pixel x_c 'sees' the shadow boundary (later referred to as the shadow time). Denote by $\Pi(t)$ the corresponding shadow plane at that time t . Assume that two portions of the shadow projected on the desk plane are visible on two given rows of the image (top and bottom rows in the figure). After extracting the shadow boundary along those rows $x_{top}(t)$ and $x_{bot}(t)$, we find two points on the shadow plane $A(t)$ and $B(t)$ by intersecting the desk plane with the optical rays $(O_c, x_{top}(t))$ and $(O_c, x_{bot}(t))$ respectively. The shadow plane $\Pi(t)$ is then inferred from the three points in space S , $A(t)$ and $B(t)$. Finally, the point P corresponding to x_c is retrieved by intersecting $\Pi(t)$ with the optical ray $(O_c, x_c(t))$. This final stage is called triangulation. Notice that the key steps in the whole scheme are: (a) estimate the shadow time $t_s(x_c)$ at every pixel x_c (temporal processing), and (b) locate the reference points $x_{top}(t)$ and $x_{bot}(t)$ at every time instant t (spatial processing).

ASSUMING:

- The intrinsic camera parameters and the desk plane location are known from pre-calibration of the camera,
- S is a point light source of known location in space (after calibration of the light source).

Both calibration steps are described in details in [1].

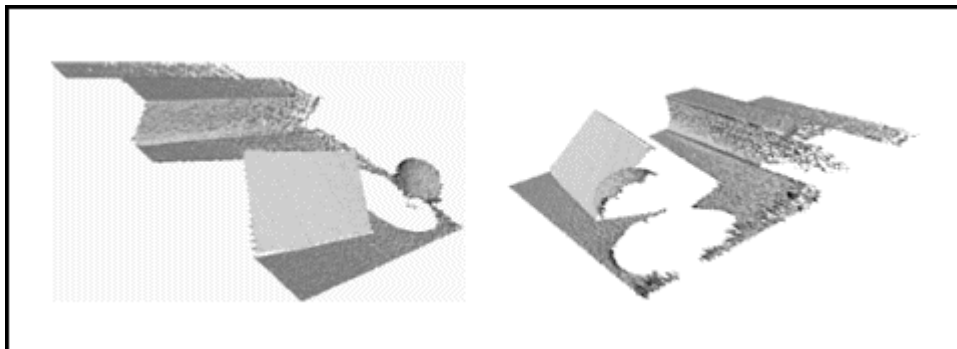
SPATIAL AND TEMPORAL PROCESSING:



The two tasks to accomplish are:

- Localize the edge of the shadow that is directly projected on the tabletop ($x_{top}(t), x_{bot}(t)$) at every time instant t (every frame), leading to the set of all shadow planes $\Pi(t)$. On the figure, the two extracted reference points are $x_{top}(t_0) = (118.42, 10)$ and $x_{bot}(t_0) = (130.6, 230)$ at frame $t_0=134$.
- Estimate the time $t_s(x_c)$ (shadow time) when the edge of the shadow passes through any given pixel x_c in the image. On the figure, the shadow time corresponding to pixel $x_c=(104, 128)$ is $t_s(x_c)=133.27$.

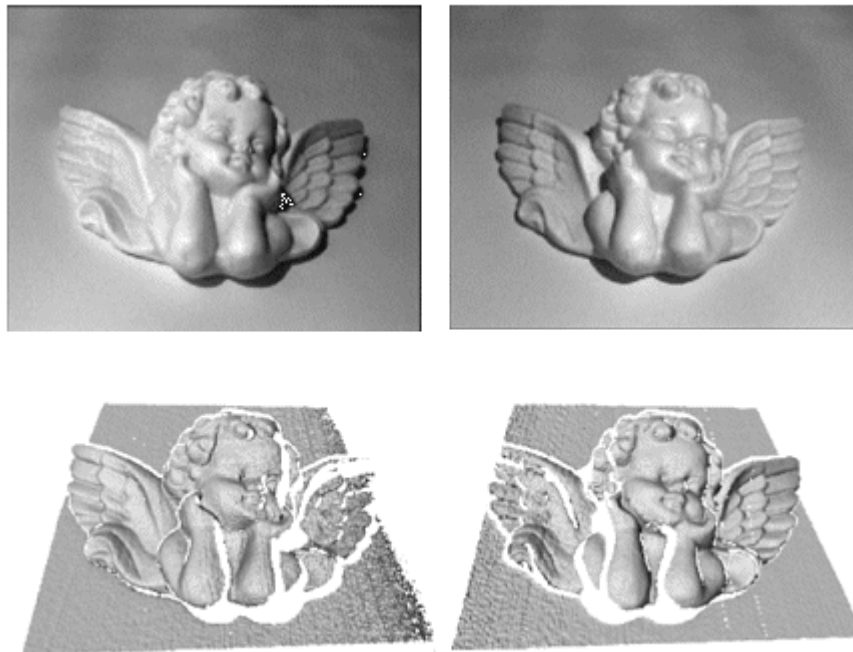
RECONSTRUCTION RESULTS:



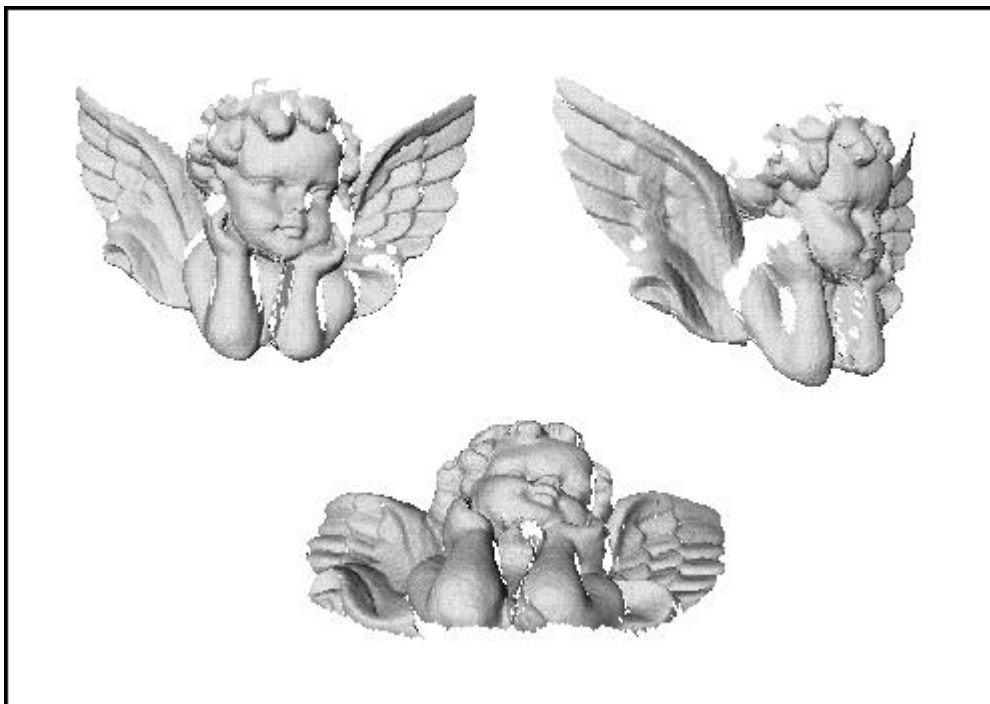
The errors of reconstruction are of the order of 0.3mm, or less than 1%

ANGEL EXPERIMENT:

Two scans of the same angel with the light source on the left and on the right of the camera:



Final 3D mesh of the angel after merging of the two scans (pixel-wise, no alignment necessary):



The errors of reconstruction are approximately 0.1mm, or 0.1% (over a depth variation of 10cm)

ACHIEVEMENTS

We have presented a simple, low cost system for extracting surface shape of objects. The method requires very little processing and image storage so that it can be implemented in real time. The accuracies we obtained on the final reconstructions are reasonable (at most 1% or 0.5mm noise error) considering the little hardware requirement. In addition, the final outcome is a dense coverage of the surface (one point in space for each pixel in the image) allowing for direct texture mapping.

PUBLICATIONS

[1] J-Y. Bouguet and P. Perona, "3D Photography on Your Desk," in Proc. of the Int. Conf. on Computer Vision, Bombay, India, January 1998. [Download postscript](#).

Patent Pending

[Back to main page](#)