3D scanning using weak structured light*

Sandeep Manandhar[1] and Ignacio Rocco[2]

***Abstract*— In this document, we present a cheap method for 3D scanning, using a LED lamp and a USB webcam. The method is based on weak structured light, i.e, the shadow produced with a paper stick when placed under the light source.**

I. INTRODUCTION

3D reconstruction is useful for a variety of purposes, including mechanical design and manufacturing quality control. Different kinds of methods exist, among which the family of methods using "structured lighting" offer the best relation between accuracy, speed and point density. However, a commercial structured light system can be costly. In this report, we present a cheap system composed of a LED light source (from a cell phone's flash light) and a USB webcam, which uses weak structured light for 3D reconstruction. It is based on the original work by J-Y. Bouguet [1].

The setup can be visualized in Fig. 1.

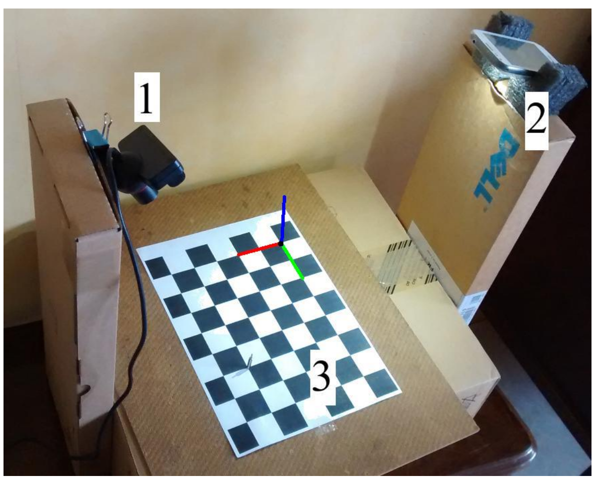


Fig. 1: Setup description: [1] USB Webcam, [2] Cell phone's LED light, [3] Work plane, and associated coordinate system

In order to obtain the 3D points using this system, a *system calibration* is needed beforehand. This will be explained in section II. Then, in section III, the geometry behind the method is presented, which allows the calculation of the 3D points by triangulation.

II. SYSTEM CALIBRATION

By system calibration, we refer to determining the relative geometry of the camera, world plane and the lamp. We assume that the intrinsic parameters K of the camera are known. During all the measurement, the relative position of these 3 elements has to remain unchanged. In order to express all this positions, we use the *world coordinate frame*, which we fix at the working plane, on the position indicated in Fig. 1. From there, the camera position C and rotation will be measured as well as the location L of the lamp, as illustrated in Fig. 2. The procedure to determine these parameters is explained in the following subsections.

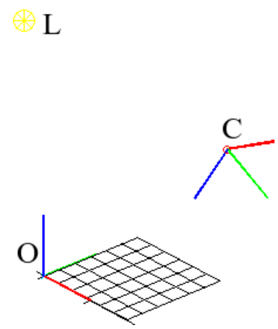


Fig. 2: The complete setup in 3d: [O] World origin, [C] Camera position, [L] Light position

A. *Extrinsic parameters of the camera*

To calculate the extrinsic parameters of the camera, namely its rotation and translation from the world origin, we first establish a homography between the checkerboard points (with known $Z = 0$) and the image formed by the camera. Four corresponding points are sufficient to get the homography matrix H. We can then extract the rotation matrix R and translation vector t from the H matrix, as indicated in the following equations.

$$p_{image} = HP_w \tag{1}$$

$$p_{image} = K[R|t]P_w \tag{2}$$

$$H = [h_1|h_2|h_3] = K[R|t] = K[r_1|r_2|t] \tag{3}$$

Eqs. 1 and 2 establish the transformation between world coordinate P_w and the image point p_{image}. We can therefore determine that the homography H can actually be decomposed into the product $K[R|t]$. In Eq. 3, the relation between H and the extrinsic rotation and translation is further studied. As in this case the world points have $Z = 0$, the third column of R can be removed from the matrix product (for further detail, see [2]). Using 4 corresponding points and the

*Based on the work by J-Y Bouguet [1]

[1]Master in Computer Vision, Université de Bourgogne, France. manandhar.sandeep at gmail.com

[2]Master in Computer Vision, Université de Bourgogne, France. rocco.ignacio at gmail.com

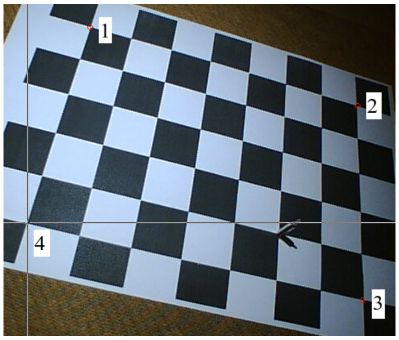


Fig. 3: Points for homography computation

direct linear transformation (DLT) algorithm, we compute the matrix H defined up to some scale λ. We then recover R and t from H. Using Eq. 3 we get: $r_1 = K^{-1}h_1/\lambda$ and $r_2 = K^{-1}h_2/\lambda$. From the cross product of r_1 and r_2, we then derive r_3. So we have a crude version of the rotation matrix. This matrix might not have the correct properties of a rotation matrix due to numerical error, and can be instead an affine transformation matrix as we can check via its determinant. To ensure its condition of rotation matrix, we fix its singular values to 1. The translation vector is given by $t = K^{-1}h_3/\lambda$. Now the camera position can be estimated as $C = -R^{-1}t$.

B. Lamp calibration

To estimate the lamp position, we use an object of known height to cast shadow in the plane $Z = 0$. Using H^{-1} we can find the world coordinates of the tip of the shadow, s and the base of the object, b by indicating their position on the acquired image. From the base location of the object, the real tip, q of the object is found. The line joining points s and q leads to the light ray l. Using multiple images of the object and its shadow in different location in the plane, we get the intercept of such lines (in the linear-least-squares sense) as the real location of the light source L. This is explained in Fig. 4.

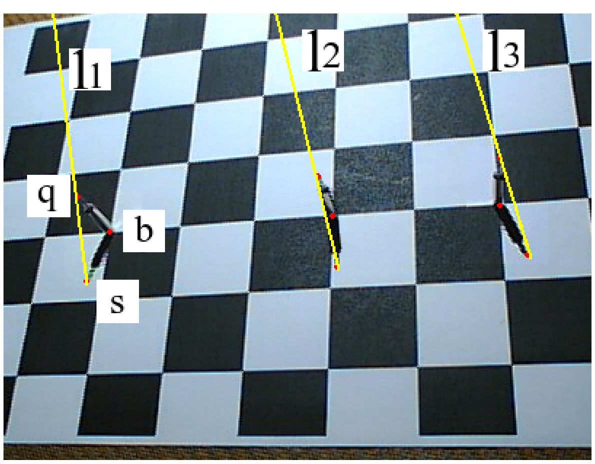


Fig. 4: Finding the light position

III. TRIANGULATION

The triangulation procedure consists in finding the 3D points of the object P_{obj} by intersecting their projection ray r with the shadow plane Π. This is illustrated in Fig. 5.

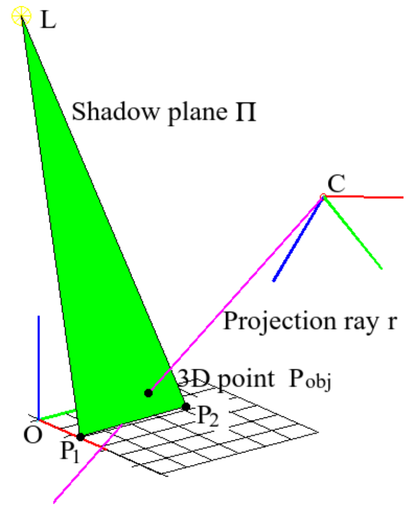


Fig. 5: Triangulation diagram

The procedure for finding the shadow plane Π and the projection ray will be explained in the following sections.

A. Shadow plane calculation

The shadow plane is determined by 3 points, the light source L and two points P_1 and P_2 on the line formed by the shadow in the work plane. The world coordinates of $P_{1\,w}$ and $P_{2\,w}$ can be determined by the image coordinates $p_{1\,image}$ and $p_{2\,image}$ by applying the inverse homography H^{-1}.

$$P_{1\,w} = H^{-1}p_{1\,image} \tag{4}$$

$$P_{2\,w} = H^{-1}p_{2\,image} \tag{5}$$

Using cross products of vectors $L\vec{P}_1$ and $L\vec{P}_2$, the normal $\vec{n}$ of the plane Π is found. Then using any of the 3 points: L, P_1 or P_2, the plane is mathematically determined as in equation 7.

$$\vec{n} = L\vec{P}_1 \times L\vec{P}_2 \tag{6}$$

$$n_1(x - x_L) + n_2(y - y_L) + n_3(z - z_L) = 0 \tag{7}$$

B. Projection ray calculation

The projection ray of a certain point $P_{obj\,w}$ in the object, and also in the shadow plane, can be determined from the image point $p_{obj\,image}$ by multiplying by $R^T K^{-1}$. This is expressed in Eq. 8.

$$u_w = R^T K^{-1} p_{obj\,image} \tag{8}$$

Where u_w is the direction vector of the the projection ray, expressed in world coordinates. The Eq. 8 shows K^{-1} takes $p_{obj\,image}$ from image plane to camera frame and eventually to world frame by R^T. The projection ray can then be expressed as:

$$r = C + \beta u \tag{9}$$

where β is just a scaling parameter. The intersection of the projection ray with the plane Π leads to the 3D point in the world frame.

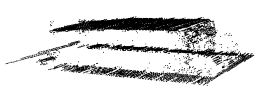
IV. RESULTS

The described algorithm was implemented in Matlab in a series of scripts. A simple GUI was also implemented to facilitate its usage.

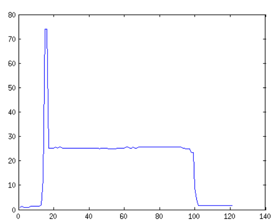
Using this tool, a series of experiments were done. The scanned scenes were the following:

- Medicine box
- Cooking spoon
- Banana
- Wool doll
- Buddha statue
- Spoon/Egg/Electric plug

The *medicine box* experiment was useful to verify the correctness of the implemented code, by checking its size against the reconstructed one. The error was estimated in about 1mm. It is to be noted that, no distortion removal has been performed.
In Fig. 6, a cross-sectional profile of the box is shown. Here, an average height of $25.1\,mm$ is measured while the real box height was $24\,mm$. For this scan, a *fast* reconstruction method was used, using only spatial processing (frame-by-frame).



(a) Reconstructed point cloud



(b) The medicine box section

Fig. 6: The medicine box

The cooking spoon, as well as the banana experiments were useful to realize the importance of self-occlusions, which lead to lost points. For these scans, and the later ones, a *dense* reconstruction method was used, by applying the concept of temporal processing, as described in [1].

Finally, a more complex scene, including an egg, a spoon and an electric plug was tried, as shown in Fig. 11.

V. CONCLUSION

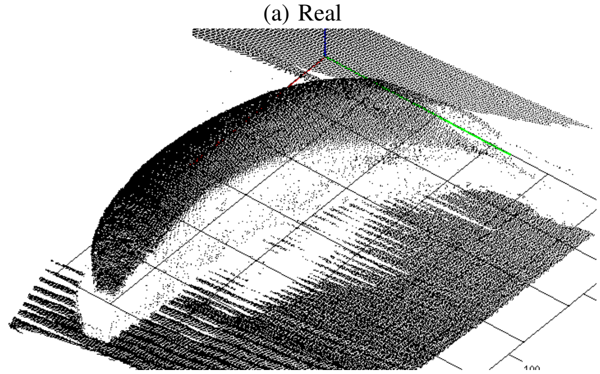
We have performed 3D scanning using weak structured light to create 3D point clouds. Using a simple setup, satisfactory results were obtained with an error around 1mm. This error could be accounted to the homography estimation as the points in the image for homography correspondence were

(a) Real

(b) Reconstructed

Fig. 7: Cooking spoon reconstruction

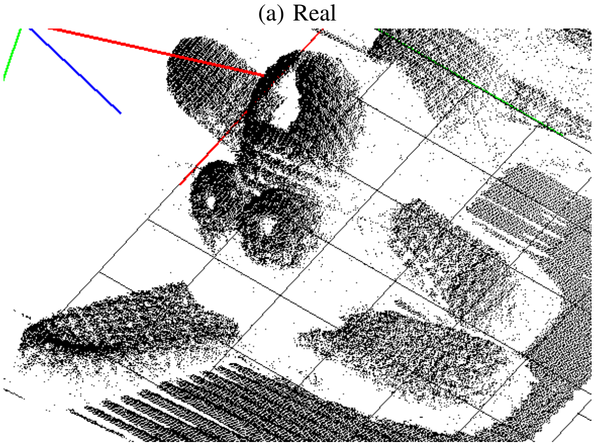
(a) Real



(b) Reconstructed

Fig. 8: Banana reconstruction

(a) Real



(b) Reconstructed

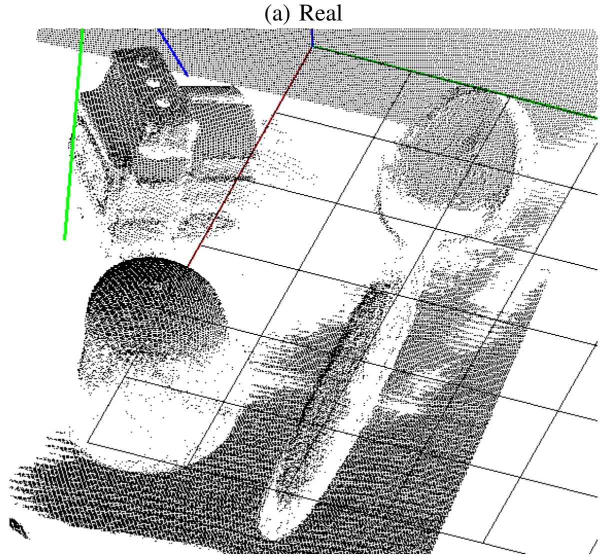
Fig. 9: Wool doll reconstruction

(a) Real Statue

(b) Reconstructed

Fig. 10: Statue of Buddha reconstruction

(a) Real



(b) Reconstructed

Fig. 11: Spoon/Egg/Electric plug reconstruction

acquired by manual clicking. Sub pixel accuracy corner-detection algorithms could help to obtain a more accurate H. Moreover, a preprocessing step, where radial distortion is removed from the acquired images could also be beneficial.

Furthermore, the methods presents one main limitation, which is the existence of missing points due to self-occlusion. However, multiple views could be used to overcome this problem and complete the lacking 3D points.

Overall, we can say that this methodology provides a cheap and fairly effective way to scan in 3D.

References

[1] J.-Y. Bouguet and P. Perona, "3d photography on your desk," in *Computer Vision, 1998. Sixth International Conference on*, pp. 43–50, IEEE, 1998.

[2] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, 2000.