SHAPE RECONSTRUCTION FROM SHADOW AND SHADING

TOH Peng Seng, GOH Wee Leng, CHAN Kap Luk

School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue
Singapore 2263

ABSTRACT

This paper presents a method of reconstructing convex polyhedral objects by the "Shape from Shadow" approach. The process utilizes several gray-scale images captured by a fixed camera with six light sources at different positions. The positions of the camera and light sources are calibrated. The two-dimensional shadows casted by an object are extracted from the images. The geometrical relationship between the light sources, the camera, and the corners of the shadow boundary are used to obtain three-dimensional coordinates of the corners of the polyhedral in a world coordinate frame. A volumetric representation of the shape of an object is then possible.

1. INTRODUCTION

Reconstructing the three-dimensional shapes of objects is one of the most important objectives in machine vision. Many approaches have been proposed. One of these methods derives the shape of an object by analyzing its shadow and is commonly known as *shape from shadow*.

Previous work on inferring information from shadows include those by Adjouadi and Tou[1], Leow *et al.*[2], and Medioni[3]. Adjouadi and Tou presented a systematic way to identify shadows from natural scenes. Leow *et al.* proposed a method to identify vertices and junctions and used them to segment the boundary of a shadow. Medioni used a simple method of obtaining the height of objects from aerial images by assuming that both the height of the object and the surface on which it is laid is parallel to each other and perpendicular to the camera axis. These three methods can be considered passive methods.

The work by Raviv *et al.*[4] can be classified as active as they used a sequence of shadows cast by moving light source to reconstruct a slice of the object. Their work assumes that the camera is far away from the object and that light rays are parallel to each other. This paper will adopt the active approach and try to remove some of the assumptions made in most other algorithms. In particular, we will tackle the problems introduced by perspective projection and non-ideal distant point source We term this the perspective approach to "Shape from Shadows".

The perspective approach aims to rectify the inaccuracy incurred from the two assumptions made in the parallel approach. The perspective approach assumes that:

1. The light rays are non-parallel, that is, the light source can be considered to be a near point source instead of a distant point source. The virtual point source position (see figure 1) can be calibrated.
2. Perspective projection from real world scene to image plane is used.

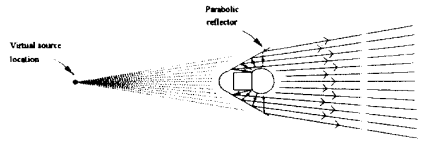


Figure 1. Characteristics of tungsten halogen lamp

A novel lighting platform known as the Ring-of-light[5] is used to provide the region illumination. The block diagram of the system is shown on Figure. 2. It consists of a ring-shape frame that houses six light sources (or any other numbers as required) pointing in the direction of the object space (which is a platform where shadow is cast). The system is controlled by a ROL controller that is further controlled by the host computer.

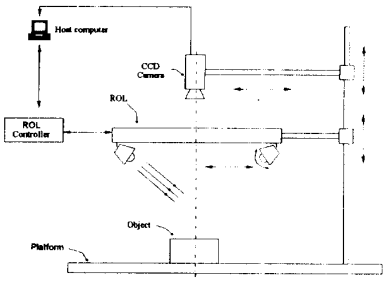


Figure. 2 Ring-of-Light Setup

Each of the light sources can be individually turned on or off as required. The direction and intensity of the light sources can be either singly or concurrently controlled. Shape reconstruction can be achieved by performing the following steps:

1. Calibration of imaging hardware,
2. Image acquisition,
3. Shadow boundary extraction,
4. Shadow boundary segmentation, and
5. Inference of height from the shadow length.

2. LIGHT SOURCE AND CAMERA CALIBRATION

Shape from shadow is a geometrical technique. The positions and the orientation of the camera and lights in a world coordinate frame must be avaiable in order to obtain true three-dimentional measurements. Hence, to find the height of an object using the shadow length, the light source's position must be determined so that the angle of elevation of the light rays can be calculated. The mapping of the shadow's absolute length on the platform to its image stored in frame buffer in pixels must be determined to convert the shadow length calculated from the image into absolute length in real world. Both the light source's position and the mapping can be determined through the calibration of a known object.

The procedure is to place a piece of graph paper onto the platform. An origin is assigned to the graph paper that will then be the origin of the coordinate system of the platform. Naturally the light source position determined will be described with respect to this coordinate frame. A rectangular object of known dimensions is placed onto this graph paper and its shadow is cast. The corners of the shadow, A' and C', are marked on the paper as shown in Figure 3.

0-7803-0891-3/93$03.00 ©1993IEEE

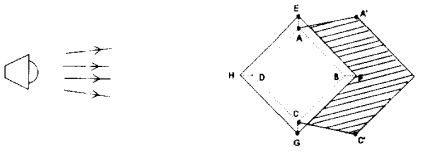


Figure 3. Shapes of shadows casted under near light sources.

Thus by drawing imaginary lines through AA' and CC', these lines will intersect at a distant point, say O, which will be the position of the light source projected onto the platform (see figure. 3). This intersection point is calculated by solving two simultaneous equations using coordinates of points A, A', C and C'. By considering the plane AA'E and the point O, the height of the light source can be calculated using similar triangles.

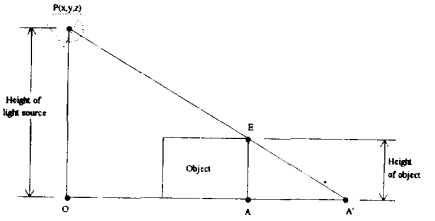


Figure 4. The height of edge EA can be obtained by similar triangle method.

$$\frac{AA'}{OA'} = \frac{AE}{AE} \qquad (1)$$

The camera is modelled as a pin-hole camera.

The optical axis intersects with the image plane forming a point known as the image centre. It has been shown by other researchers[6] that this is not always true. The image centre and hence the position of the optical axis can be estimated by calibration.

A square rectangular box is drawn on a piece of white paper and place onto the platform. The box is captured. The coordinates of the four corners, i.e. (x,y), where x and y are its column and row number respectively, are recorded down. By changing the focal length of lens, the image of the box is enlarged or reduced. The coordinates of the four corners are again recorded. The procedure is repeated to generate enough readings. These readings are then reproduced on another piece of graph paper and four best-fitting straight lines are drawn joining the corners as shown on Figure. 5. The coordinate of the point where these four best-fitting lines meet is the estimated optical centre in the captured image. Knowing this optical centre, the exact position of the optical axis can be located on the platform in terms of the platform's coordinate frame.

The mapping of CCD sensor's area to frame buffere is not always one-to-one. Scaling factors are therefore involved in these mapping. Figure. 6 shows a rectangular box sensed by a CCD camera and its image stored in the frame buffer.

Since the object is not at infinity, the relationship between u, v and f is given by the lens equation:

$$\frac{1}{f} = \frac{1}{-} + \frac{1}{v} \qquad (2)$$

where u is the object to lens distance, v is the distance from optical centre to image plane and f is the focal length of the lens. By similar triangle,

$$\frac{y}{v} = \frac{t}{v} = \frac{t'/k_r}{v} \qquad (3)$$

$$\frac{x}{v} = \frac{s}{v} = \frac{s'/k_c}{v} \qquad (4)$$

where k_r, k_c are the scaling factors row-wise and column-wise respectively.

The scaling factors that map photosites onto the frame buffer is given by:

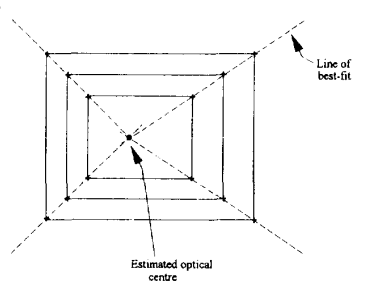


Fig. 5. Locating the optical centre of lens.

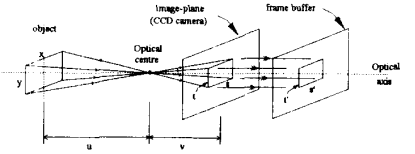


Figure. 6. Coordinate system and planes.

$$k_c = \frac{s'u}{xv} \qquad (5)$$

$$k_r = \frac{t'u}{yv} \qquad (6)$$

In determining the absolute shadow length, the main concern is the relationships between y and t', x and s'. They are :

$$\frac{t'}{y} = \frac{k_r v}{} = S_r \qquad \text{pixels/ unit length} \quad (7)$$

$$\frac{s'}{x} = \frac{k_c v}{} = S_c \qquad \text{pixels/ unit length} \quad (8)$$

where S_r and S_c are the overall scaling factors which are dependent on v and u. If the focal length, f, and the object-to-lens distance, u, remain constant, then the scaling factors, S_c and S_r, which map absolute distance into number of pixels in the memory buffer, can be determined by calibration without having to know the true values of k_c, k_r, f, v, and u.

2.4. SHADOW BOUNDARY SEGMENTATION

Separating the shadow from the background requires selecting a threshold value above which a pixel will be classified as white and otherwise, black. The threshold level is critical since it can cause overall shrinkage of the shadow. Otsu's threshold selection method[6] is used. Otsu's automatic selection of threshold value is based on the grey-level histogram of the image and uses measures derived from discriminant analysis which maximizes the separation of class mean and minimizes within class variance. The method automatically selects the optimal threshold value by evaluating the 'goodness' of the threshold value and does not require *a priori* knowledge of the scene.

Contour following[7] technique is used to extract shadow boundary. The boundary segment is chain-coded. The chain-code is also modified to facilitate corner detection. Each point on a shadow boundary is either an entry or exit point[9]. Segments consisting of entry-points are called entry-segments; segments of exit-points are called exit-segments. The exit-segments are further away from the light source than the entry-segments(see Figure 7).

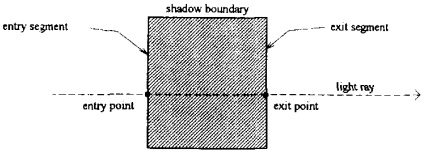


Figure 7. The light ray enters and exits a shadow at its boundary.

Corners of the shadow captured are needed to indicate the starting and ending pixels of both entry-segment and exit-segment. Whether a segment is entry or exit depends on the position of the light source. While exit-segment is the boundary of the shadow region, entry-segment is part of the object's boundary and it contains information about the location of the object. The position of all pixels along entry-segment thus must be preserved. After the corners have been found, they are classified into entry-corners or exit-corners.

After thresholding, the binary image of the shadow is as shown in Figure 8. The background is white and the shadow is black. We have implemented a simple technique to identify entry-corners from all edge corners.

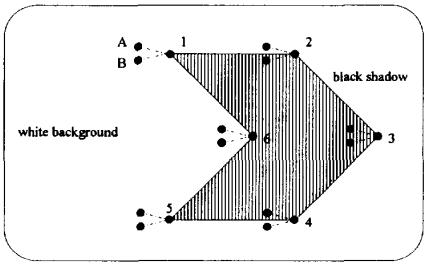


Figure 8. Plan view of shadow region after thresholding.

To classify a pixel as an entry-pixel, one need only to check the intensity values of its neighbour pixels. The pixels between a light source(in this case, parallel light rays) and an entry-pixel should have intensity value of white. In particular, If any two pixels A and B a few pixels away from the suspected corner pixel have intensity white, then that pixel is an entry-corner pixel. In Fig. 8, pixels 1, 5 and 6 are thus entry pixels while pixels 2, 3 and 4 are not. Note that although pixel 6 is also selected, it is redundant and can be discarded since pixels 1 and 5 are enough to define the start pixel and the end pixel of the entry-segment. Thus if the classification of the corners occurs in a clockwise direction starting from pixel 1, then pixel 1 will be the first to be put into a list of entry-corners, followed by pixel 5 and lastly pixel 6. To identify the entry-segment, one need only to extract the (x,y) coordinates of the first entry-corner and the last entry-corner from the list of entry-corners. In this way, pixel 6 is naturally discarded.

The selection of the locations of the two test pixels A and B is crucial. If they are too far apart, pixel 3 will be considered an entry-corner while pixel 6 will not be an entry-corner. If they are too near each other, it approaches the case of using just one test pixel. In such a case, one cannot tell whether pixels 2 and 4 are entry- or exit-pixels as the test pixel lies exactly on the boundary. Thus setting a separation of a few pixels between pixels A and B will be sufficient for the task.

4. INFERENCE OF HEIGHT BY PERSPECTIVE APPROACH

Consider an ideal pin-hole at a fixed distance in front of an image plane. Assume that an enclosure is provided so that only light coming through the pin-hole can reach the image plane. In perspective projection, each point in the image plane corresponds to a particular direction defined by a ray from that point through the pin-hole. This direction is also known as the line-of-sight vector of the image point [8]. Figure 9(a) illustrates how the image of a box is formed on the image plane under perspective projection and Figure 9(b) shows what the box will look like on the screen.

Due to perspective projection, points on planes that are nearer to the pin-hole will be projected onto points on the image plane further away from the centre of the image plane. Hence for a rectangular object placed directly under the camera, what appears on the image plane is the object's top surface and its based boundary is hidden as illustrated in Fig. 10(a) and (b).

Fig. 11 shows how the object's shadow will be captured and displayed on the display screen. Because the object's base EFGH is hidden, what is captured and shown on the screen is the shadow region F"F'G'G" instead of FF'G'G. If the segment F"G" is taken as the object's base boundary and is

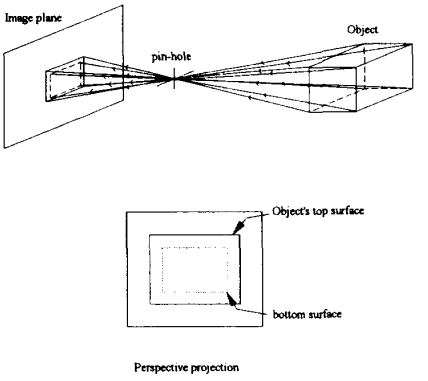


Figure 9. Perspective projection of three-dimensional object onto image plane.

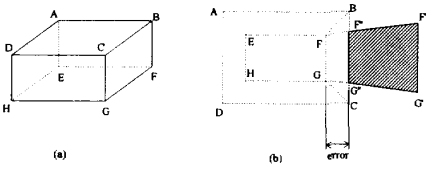


Figure 10. Presence of perspective error.

back projected into the real world coordinate frame, then error occurs. Ironically, this error is proportional to the object's height. Because of this error, the object's exact position on the platform and the height of the object cannot be determined accurately.

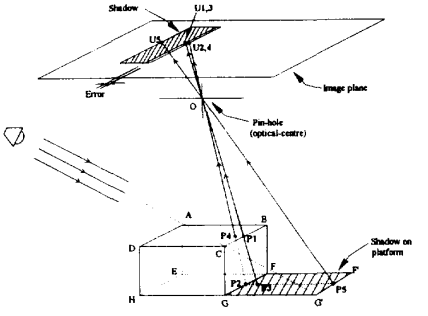


Figure 11. Back projection of shadow image points onto real world coordinates.

The following section describes a method that will be able to determine the object's exact location and its height in the real world accurately. The intention of this approach is to look for a way of determining the plane ABCD in Fig. 11 onto which the image point U1 can be back projected as P1. Once the coordinate of point P1 is known, the object's base boundary is just a projection(vector dot product) of point P1 onto the platform.

Using this method, the shadow length or width on the image plane is not used at all and it is more realistic as it does not have to assume the light source is at infinity and that the light rays are parallel. However, there are two assumptions made. First, it is assumed that the lens aperture is small so that the camera can be modelled as a pin-hole camera. Second, the pin-hole is assumed to be at the aperture ring of the lens so that its distance from the platform can be physically measured using a measuring ruler.

4.1. COORDINATE SYSTEM CONVENTION

The coordinate assignment is shown in Fig. 12. For ease of presentation, the image plane is shown placed in front of the pin-hole so that the image formed is not inverted.

The imaging system illustrated in Fig. 12 uses the following conventions:

- All origins are aligned with the optical axis and all z-axis parallel with the optical axis.
- The ground plane (platform) is taken as the reference plane.
- The camera points towards the negative z axis direction.
- The camera is modelled as a pin-hole camera with the pin-hole(optical centre) assumed to at the aperture ring at a distance u from the ground plane.
- The image plane is parallel to the ground plane and at a distance v from the pin-hole, therefore the camera is pointing straight down to the centre of the ground plane, noted as O' (Origin).
- The TV image is subdivided into a rectangular grid having 512 horizontal lines with each line containing 512 adjacent pixels. Each pixel represents one digitized value. This convention is used by the frame grabber. The origin is at the point where the optical axis passes through the image plane. Therefore conversion has to be performed when viewing a pixel on the monitor to a point corresponding to the image plane.

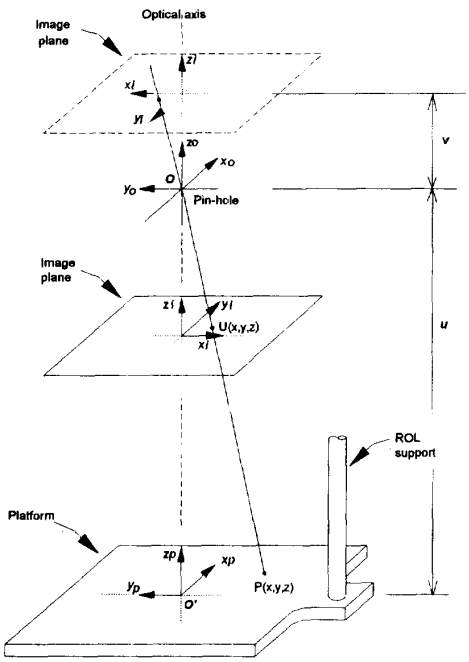


Fig. 12. Coordinate frame assignment.

4.2. GENERAL PROCEDURE OF PERSPECTIVE APPROACH

In using this approach, the following conditions have to be satisfied.

- The light source position (its coordinates) with reference to the platform's coordinate frame is known. This can be done by calibration mentioned in section 2.1.1.
- The optical axis and optical centre's (pin-hole) coordinate in platform's coordinates frame are known. The optical axis can be estimated as described in section 2.1.2 and the optical centre position can be found by measuring the height of the lens's aperture ring from the platform.
- The scaling factors, S_c and S_r, have been determined through calibration for a given focal length and object-to-lens distance (i.e. u). Once the focal length or object-to-lens distance is changed, S_c and S_r have to be re-calibrated.

Fig. 13 shows the ring-of-light with one of the light bulb(bulb A) switched on. The coordinates of the light bulb and optical centre are as indicated.

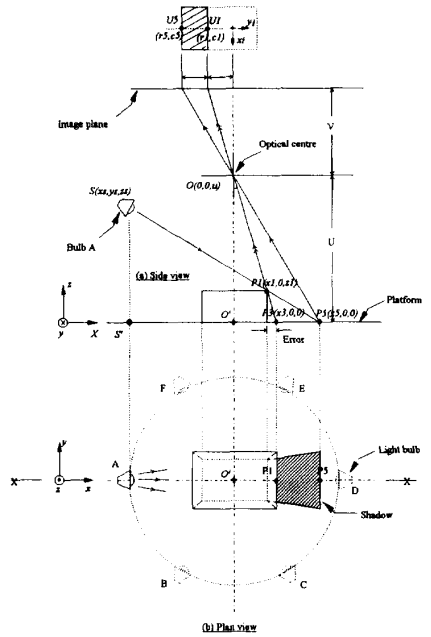


Fig.13. Perspective projection of shadow onto image plane.

The first step is to find two points, U1 and U5, on the shadow region on the image plane, so that after multiplying their coordinates(in image plane's coordinate frame) by the scaling factors, they are back projected onto points P3 and P5 which will form a plane together with the light bulb and origin of the platform's coordinate frame, point P3 and P5. The algorithm used to locate these two points, U1 and U5 will be explained shortly.

With P3, P5, the light source position and the optical centre all lie in the same plane, finding the coordinates if P1 is now reduced to a two-dimensional problem with the calculations simplified. P1 can be found by simply solving two simultaneous equations. They are the line equations of SP5 and OP1. With reference to Fig. 13, the coordinates of P3 and P5 are obtained as follow :

$$x_5 = \frac{c_5}{S_c} \tag{14}$$

$$y_5 = \frac{-r_5}{S_r} = 0 \tag{15}$$

$$z_5 = 0 \tag{16}$$

$$x_3 = \frac{c_1}{S_c} \tag{17}$$

$$y_3 = \frac{-r_1}{S_r} = 0 \tag{18}$$

$$z_3 = 0 \tag{19}$$

where r and c are in terms of pixels. The above expressions are equivalent to multiplying the point (r,c) with the transformation matrix given by:

$$T = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{20}$$

By using the x and z coordinates of the light source, optical centre, point P3 and P5, two simultaneous equations can be formulated. The solution of this set of equations gives the x and z coordinates of point P1. The y coordinate is zero for this particular light source position.

Since the Ring-of-light consists of six light bulbs, by switching off bulb A and switch on the next one, say bulb B, another point on the object can be calculated. This is illustrated in Fig. 14. Point P8 and P10 are points on the platform while P6 is on the object.

Light bulb B is at an angle of 60° away from bulb A so the coordinates of Bulb B can be easily derived from A. The procedure is similar to when bulb A is on. The only difference is that the plane consists of the light bulb B, point P8 and P10 are not the x-z plane of the platform's coordinate system. To simplify the calculation, this plane is still used to solve the two simultaneous equations of line SP10 and OP8. However the solution of the two equations will only give point P6's z coordinate (i.e. its height) and its absolute distance from the platform origin. The x and y coordinates of P6 can be found using simple trigonometry to decompose this absolute distance into x and y components. Therefore by switching on the light bulbs from A to F, six points from the object can be calculated. This obviously is the limitation of the approach as it is impossible to deduce the entire shape of the object from these six points. If more light bulbs are available, then more points can be 'sampled' and calculated. Then the whole object shape can be deduced by method of interpolation to generate the remaining points on the object. On the other hand, if instead of having more light bulbs, there is only one light bulb but can be made to maneuver 360° around the object then again more points can be calculated.

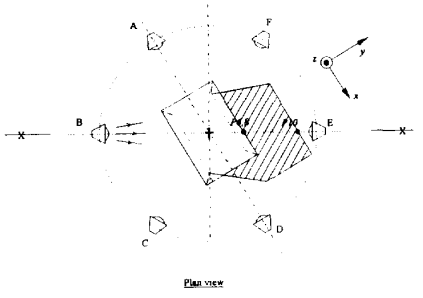


Fig. 14. Calculation of height.

The main advantage of using this approach is that the problem is broken down into solving simultaneous equations in two-dimensional plane instead of solving complicated plane and line equations in three-dimensional space.

4.3. PROCEDURE FOR FINDING INTERSECTION OF LINES

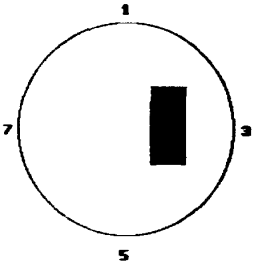
Beside calculating the exact points of the object in real world, the essential part of the routine is the search for the corresponding points on the entry- and exit-segments, points U1 and U5 on Fig. 14 for example, for each particular light source position.

The shadow on the screen has already been chain-coded and hence for all the pixels that form the shadow boundary, each pixel's row and column numbers are known. With the light source position known after calibration, the projection of light source's position onto the platform can be determined, let it be S'. The routine then calculates the gradient of the line joining point S' and the origin of the platform O'(see Fig. 14). Let this gradient be GO. Next the routine scans the chain-codes pixel by pixel. Each pixel's row and column numbers are converted to the image's coordinate system and multiplied by the scaling factors to back project them onto the platform, say P5 for instance. The gradient of the line joining P5 and O', GP, is calculated and compared with GO. If their difference is not within a given tolerance, P5S'O' is obviously not collinear and P5 is discarded. The routine proceeds to select the next pixel on the chain-codes until the difference between GO and GP is within the tolerance. Either along the entry or exit-segment, there bound to be more than one pixel which will meet the tolerance. Out of these pixels, on say exit-segment, the routine selects the pixel which best meets the tolerance as the most suitable choice for P5. In other words, S'O'P5 will be collinear. Similarly, another point, P3 is chosen on the entry-segment. The routine does not have to differentiate between exit- or entry-segment. By having a tight tolerance, one complete search over the chain-codes will definitely produce only two points, P5 and P3 from the two segments. With P5P3OS' forming a plane, the routine then calculates point P1 by solving two simultaneous equations(lines OP3, SP5). When the light source is switched to another position, the shadow is chain-coded again and the routine repeats. When the light source switches from A to B then finally to F, six points on the object are calculated.

5. RESULTS AND DISCUSSIONS

The illumination were supplied by high intensity tungsten halogen lamps integrated with specular mirrors having faceted surfaces. The pictures were taken by a CCD camera through a 25 to 75 mm zoom lens. Ambient light are cut out by enclosing the whole setup in a 1m x 0.75m x 1m wooden black box. The target object used is a simple rectangular block 35 x 25 x 20 mm in size.

The following section shows the results of using the perspective approach (see Photograph 1).



Photograph 1 Shadow cast by a rectangular block

In the perspective approach, only six points are calculated. The results are shown on Fig. 15. (Xh, Yh, Zh) give the coordinates of each point in the platform's coordinate system. Zh is thus the height of the object.

i	Xh (mm)	Yh (mm)	Zh (mm)	Perspective Error (%)	Height Error (%)
0	15.621	0.000	19.717	3.155	1.414
1	7.076	12.256	20.522	3.284	2.611
2	-7.002	12.128	20.170	3.227	0.852
3	-16.480	0.000	22.592	3.615	12.960
4	-6.846	-11.858	18.814	3.010	5.932
5	6.938	-12.018	17.447	2.792	12.764

Fig. 15. Results of the six points calculated from perspective approach.

The correct values of these coordinates are calculated as shown on Fig. 15 and they are tabulated on Table 1.

The results obtained using perspective approach on another object, a 5-sided polygon, is shown below. Photograph 2 shows the detecting and locating of the entry and exit points from the shadow at a particular light bulb position.

Fig. 17 shows the actual coordinates of the 6 points on the object. Fig. 18 shows the tabulated coordinates of the six points determined by the program for one complete round.

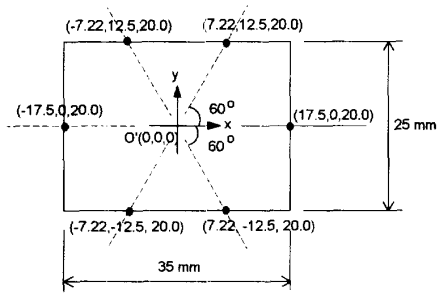


Fig. 16. Correct coordinates of the six points.

i	Xh	Yh	Zh
0	17.5	0	20.0
1	7.22	12.5	20.0
2	-7.22	12.5	20.0
3	-17.5	0	20.0
4	-7.22	-12.5	20.0
5	7.22	-12.5	20.0

Table1. Table of correct coordinates.

The percentage errors are calculated and shown on Table 2.

Xh (%)	*Yh* (%)	*Zh* (%)
10.74	0	1.41
1.99	1.95	2.61
3.02	2.98	0.85
5.83	0	12.96
5.18	5.14	5.93
3.91	3.86	12.76

Table 2 Percentage error of calculation.

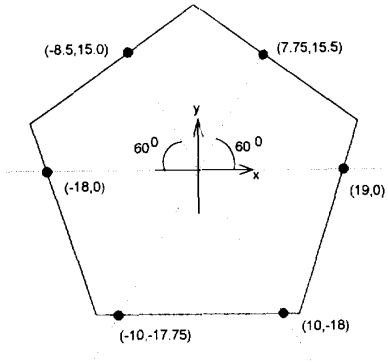
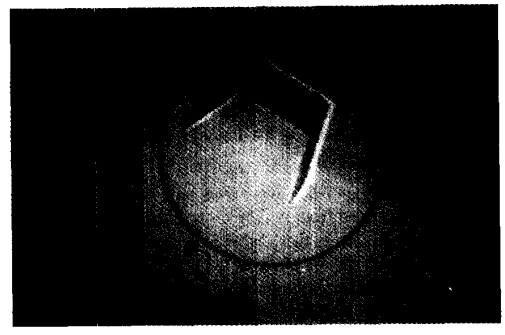


Fig. 17. Actual coordinates of the 6 points on the polygonal object.

i	Xh (mm)	Yh (mm)	Zh (mm)	Height Error (%)
0	17.793	0.000	20.229	1.143
1	7.343	12.718	20.504	2.519
2	-6.927	11.998	20.166	0.830
3	-16.215	0.000	20.865	4.323
4	-10.246	-17.747	19.600	2.002
5	10.619	-18.394	17.940	10.302

Fig. 18. Tabulated results of the 6 points found by the program.



Photograph 2 Shadow cast by a 5-sided polygon

CONCLUSIONS

A method of reconstruction the shape of convex object has been described. The perspective approach, aimed to improve on the parallel approach by assuming perspective projection and non-parallel light rays from a point source. Shape from shadow method for determining object shape depends on a detailed understanding of the imaging process. The experiment has demonstrated that shape reconstruction from shadows is feasible. The results from both approaches were obtained using a simple rectangular object. With a typical field of view of 20 mm, the perspective approach had obtained an average of 5% error in calculating the height and position of the object respectively. Errors in the calibration procedures need to be investigated and overcome. It is realistic to expect future advances both in the calibration method and the hardware to achieve the necessary accuracy.

ACKNOWLEDGEMENT

The authors would like to thank Lim Lian Swa and Lim Lye Kheng for carry out this work.

REFERENCES

[1] M. Adjouadi and J.T. Tou, **Shadow Analysis in Scene Interpretation**, *Image Analysis, Proceedings of the 4th Scandinavian Conference*, Trondhein, Norway, Vol. 2, June 17 - 20, 1985, pp 821 - 829.

[2] M.H. Loew, L.N. Hambrick, and R.L. Carroll, Jr., **The Entry-Exit Method of Shadow Boundary Segmentation**, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, September 1987.

[3] G.G. Medioni, **Obtaining 3D information from Shadows in Aerial Images**, *Proceedings of Computer Vision and Pattern Recognition Conference*, Washington, DC, 1983, pp 73 - 76.

[4] D. Raviv, Y.H. Pao and K.A. Laparo, **Reconstruction of Three-Dimensional Surfaces from Two-Dimensional Binary Images**, *IEEE Transactions on Robotics and Automation*, Vol. 5, No. 5, October 1989, pp 701 - 710.

[5] P.S. Toh, W S. Ching, **Automatic Optimization of Machine Vision Lighting**, *Proceeding ,of International Conference on Image Processing, Singapore, 1992.*

[6] Lenz R.K and Tsai R, **Techniques for calibration of the scale factor and image centre for high accuracy 3-D machine vision metrology**, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, September 1988.

[7] P.K. Sahoo, S. Soltani, A.K.C. Wong and Y.C. Chen, **A Survey of Thresholding Techniques**, *Computer Vision, Graphics and Image Processing*, Vol. 41, 1988, pp. 233-260.

[8] R.C. Gonzalez, P. Wintz, **Digital Image Processing**, Addison Wesley, 1987, pp. 297 - 233.