Six-dimensional vision system

J. Albus, E. Kent, M. Nashman, P. Mansbach, L. Palombo
National Bureau of Standards, Washington, D.C. 20234

M. Shneier
Computer Science Center, University of Maryland
College Park, Maryland 20742

Abstract

There are six degrees of freedom that define the position and orientation of any object relative to a robot gripper. All six need to be determined for the robot to grasp the object in a uniquely specified manner. A robot vision system under development at the National Bureau of Standards is designed to measure all six of these degrees of freedom using two frames of video data taken sequentially from the same camera position. The system employs structured light techniques; in the first frame, the scene is illuminated by two parallel planes of light, and in the second frame by a point source of light.

Introduction

The goal of the present research is to design a simple structured light robot vision system suitable for visual servoing in real time. This implies the capability to completely specify the position and orientation of a surface relative to a robot end-effector in real time. The definition of the position and orientation in space of an object relative to a robot gripper has six degrees of freedom. All six need to be determined to enable the robot to grasp the object in a uniquely specified manner. The system presented here has this capacity with planar surfaces, but the principle is extensible to curved surfaces as well. The present system is designed for servoing but may be extended in the future to perform object or surface recognition. For these functions, prior knowledge in the form of models and additional vision processing will be needed.

The use of structured light to extract three-dimensional shape information is a well-known technique. The use of single planes of light to measure the distance and orientation of objects has been reported several times. [1,2,3] The use of multiple planes of light [4,6,9] and multiple points of light [5,7] to measure shapes and construct volumetric estimates of objects has also been widely reported.

In our system, both the structured light projector and the camera are mounted on the robot hand in the configuration shown in Figure 1. Two frames of video data are used. The first frame is taken while illuminating the scene with two parallel planes of light provided by two slit projectors. In the second frame, the same scene is illuminated by a point source flood light located between the slit projectors.

The first image, which is produced by the pair of planes of light, makes it possible to compute the range, and the pitch and yaw orientations, of any simple geometrical surface. The second image, produced by the point light source, can then be analyzed to obtain the elevation, azimuth, and the roll orientation of the illuminated surface. The combined information then uniquely determines the six degrees of freedom of the surface relative to the vision system. When performed on binary run-length encoded images, this analysis can by performed quickly enough to be used for six-axis visual servoing.

Analysis of the Plane of Light Images

The basic geometrical relationships are illustrated in Figure 1. The vision coordinate system is so defined that the optical center of the camera lens is located on the z-axis at $(0,0,z_c)$, and the visual axis of the camera lies in the y-z plane. Near the origin, a pair of slit projectors is located in such a manner that the plane of light from the lower slit is coincident with the x-y plane. The entire vision system is mounted on the robot hand in such a manner that the vision y-axis is parallel to the roll axis of the robot hand, and the vision x-axis is parallel to a line through the two robot fingertips. The upper plane of light is parallel to the lower with a distance between them of z_t.

In Figure 1, a plane surface of arbitrary shape is shown illuminated by the two planes of light. The lower slit projector produces a light stripe across the surface from bl through bm to br. The upper slit projector produces a light stripe across the surface from tl through tm to tr. The points bm and tm lie in the y-z plane. Figure 2 is a view of the surface as seen by the camera.

The y coordinate of any pixel on the line bl:bm:br can be found directly from Table 1 by formula (1),

$$y = z_c \tan\gamma_b \tag{1}$$

and the x coordinate of the same pixel can be found using Table 1 and formula (2), where j_b is the pixel row number (as defined in formulae (A1)-(A6) in the Appendix).

$$x = i_b \left(\frac{s\, r'}{f'} \right) \tag{2}$$

The contents of Table 1 are generated by formulae (A1)-(A6) in the Appendix.

The z coordinates of all points on the bl:bm:br line equal zero, because the bottom light stripe lies entirely in the z=0 plane.

The x and y coordinates of any point on the line tl:tm:tr can be found by a similar procedure from Table 2. This table is constructed by formulae (A7)-(A12) in the Appendix. The z coordinate of any point on the line tl:tm:tr equals z_t because the top plane of light lies in the $z=z_t$ plane.

The tangent of the angle α by which the plane of the surface is rotated with respect to the x-axis can now be found from

$$\tan\alpha = \frac{y_{br} - y_{bl}}{x_{br} - x_{bl}} \tag{3}$$

where x_{br} and x_{bl} are the x-coordinates and y_{br} and y_{bl} are the y-coordinates of the points br and bl.

The points bm and tm of Figures 1 and 2 are defined by the intersection of the images of the bottom and top light stripes with the central vertical column of pixels in the camera frame. Thus, the x-coordinates of both bm and tm are zero. The y-coordinate of bm and tm can be found from Tables 1 and 2.

The tangent of the angle β by which the plane of the surface is rotated with respect to the z-axis can now be found from

$$\tan\beta = \frac{y_{tm} - y_{bm}}{z_t} \tag{4}$$

The equation of the plane of the surface may now be determined, and is given by

$$y = y_{bm} + x \tan\alpha + z \tan\beta \tag{5}$$

The first camera image, obtained with the two planes of light as structured illumination, thus provides information as to three of the six degrees of freedom of the plane of a surface with respect to the camera.

Analysis of the Point Illumination Image

In order to determine the remaining three degrees of freedom, a second frame is obtained using a point illumination source located at or near the origin. The point source illumination produces an image of the entire surface which is thresholded and analyzed by standard techniques of connected region analysis. This reveals feature points such as edges, corners, holes, and the centroid and principal axis of the image (8) The coordinate values of any point in this second image can be computed by employing a third calibration procedure, which allows us to write the equation of the ray defined by any pixel.

Any point on the ray defined by a pixel at row jp and column ip has a z-coordinate which can be computed from formula (6) and Table 3. The x-coordinate which can be computed from formula (7). See Figure 5.

(6) $$z = z_c - y \cot \gamma_p$$

(7) $$x = i_p \cdot y \left(\frac{s}{f' \sin \gamma_p} \right)$$

procedures for obtaining Table 3 are outlined in the Appendix.

Equations (5), (6), and (7) can now be solved to compute the x,y,z coordinates of any point on the surface of the object.

Obtaining Roll Orientation

Any two points on the surface can be selected to define the roll orientation of the object. For example, two points along an edge, or points identifying two features such as holes might be chosen. The slope of the line connecting these two feature points in the plane of the object surface defines the roll orientation of the object. For approaching and grasping the object there typically will exist some desired roll orientation of this line. The angle between the desired line and the measured line is then the roll angle which is required to bring the measured line into alignment with the desired line. α and β computed in formulae (3) and (4), correspond to yaw and pitch respectively.

Alternatively, the x,y,z, coordinates of any three points corresponding to three recognizable non-colinear features on the object can be computed, and then used to uniquely determine the six degrees of freedom of the object relative to the coordinate system of the robot hand in which the camera is fixed.

Dealing with Solid Objects

A problem often encountered in dealing with three-dimensional objects is that more than one surface may be visible to the camera. If the lines produced by the two planes of light touch more than one surface, this will be apparent as corners in the lines (where differently oriented surfaces meet) or as breaks in the lines (at surface depth discontinuities); the line segments between these events are simply treated as defining different surfaces.

In some cases, the lines produced by the two planes of light may lie entirely on one face of the object (evidenced by the fact that they are straight and parallel) and yet all the corners visible in the floodlit image may not lie on the same surface as the lines. This could lead to errors in computing the positions of edges and corners.

In the case of convex rectilinear objects, this difficulty can often be overcome by a simple set of rules. For example, if the two lines provide the equation of the surface that reflects them, and this equation indicates the surface slopes away to the left and away at the bottom, then points on the left and bottom edges of the flood image will be on the same surface as the lines. Thus the left and bottom edges can be used to calculate the position and roll orientation of the surface. The top and right edges are not necessarily on the computed surface and should not be used for further computations. With more complex objects, more complex rules and *a priori* knowledge about the object is required to resolve ambiguities.

Experimental Data

Photographs 1 through 4 illustrate the operation of the system, and show some of the problems encountered with an object of complex shape. The object, shown in normal view in Photograph 1, is a picture of a shiny metal object consisting of a flat rectangular plate with a rectangular block standing in the lower left and a cylinder standin gin the upper right corner. The view point of the camera, which is diagrammed in Figure 3, makes the object appear to be upside down. This is due to the fact that the camera is overhead looking down and back on the object, and the light source is located underneath the camera. There are tool marks on the surface resulting from the machining process that produced the object. These reflect the light in complex ways. Photograph 2 is the thresholded binary image obtained from the pair of planes of light, and Photograph 3 from the floodlight. Photograph 4 shows a superimposition of the two images from Photographs 2 and 3.

One notable feature of these images is the fact that despite the shiny surface with machining marks, the images are easily processed. The line images have a few extraneous points due to highlights from scratches, but these are easily removed. The break in the bottom line illustrates the effect of a step in the surface. The straightness of all the line images indicates that they lie on flat surfaces. The termination of the right end of the upper line is caused by its occlusion by the cylinder. The termination of both lines before reaching the left edge of the image is a result of their rapid fall-off in intensity due to the high reflectance of the shiny surface.

It is clear that these images are sufficiently ambiguous that the shape of the objects cannot be determined without further information. However, this system is intended for determining orientation where shape is known, as in a machine shop, where a priori knowledge will be available in the form of models of the objects from a computer-aided-design database, or from sample parts. The next step in the development of this vision system will be the use of models to remove the ambiguities in the observed image data. In this example, the pair of lines at the left indicate one surface orientation; the pair at the right indicate another. The model can be tried first with one of these orientations and then the other. If the positions of the top and bottom edges of the model are successfully fitted to those of the floodlit image, the ambiguity is resolved. Only a small number of ways of fitting are possible.

Once the robot has moved to a position such that the y-axis of the vision coordinate system is normal to the object surface and near the centroid of the surface, then the top edge will cast a shadow in the point source image. Both sides of the object surface will also be discernable from the two line image as well as from the point source image. Thus, if the vision system is continuously recomputing 6-D information as it approaches a target object, the robot hand will servo to the correct grip position regardless of the original position and orientation of the object relative to the camera.

Analysis has so far been done only with rectangular objects. However, it appears that if the shapes of the target objects are known to the robot, it will be possible to extend this analysis to the other geometrical shapes, particularly where the expected shape of the intersection of the planes of light with the surface is easily computable (e.g., cylindrical surfaces). The integration of this 6-D vision system with internal models of objects is a research topic which is being actively pursued.

Appendix

The first column of Table 1 can be calculated from formulae (A1) through (A3) which can be derived by inspection of Figure 4.

(A1) $$y = z_c \tan \gamma_b$$

(A2) $$\gamma_b = 90^\circ - A_{12} - \phi$$

(A3) $$A_{12} = -\tan^{-1}\left(\frac{j_b \cdot t}{f}\right)$$

The computation of the value of x is accomplished

(A4) $$x = i_b \cdot \left(s \frac{r'}{f'}\right)$$

The second column of Table 1 consists of the coefficient of i_b in formulae (A4)

(A5) $$r' = \sqrt{z_c^2 + y^2}$$

(A6) $$f' = \sqrt{f^2 + (j_b \cdot t)^2}$$

t = pixel row spacing on chip

s = pixel column spacing on chip

j_b = pixel row in image of bottom line

i_b = pixel column in image of bottom line

f = distance of chip from center of lens

ϕ = angle of inclination of camera axis with y-axis

The value of t, s, f, and ϕ can be determined empirically by measuring the pixel row and column of a few points in the x-y plane and performing a least squares fit for t, s, f, and ϕ. Tables 1-3 are constructed using the values z_c = 20 cm, z_t = 2 cm, f = 0.9 cm, t = .0046 cm, s = .0036 cm, $\phi = 43.2^\circ$.

Table 2 can be derived in a similar fashion from formulae (A7)-(A12).

(A7) $$y = (z_c - z_t) \tan \gamma_t$$

(A8) $$\gamma_t = 90^\circ - A_{12} - \phi$$

(A9) $$A_{12} = -\tan^{-1}\left(\frac{j_t \cdot t}{f}\right)$$

(A10) $$x = i_t \cdot \left(s \frac{r''}{f''}\right)$$

where

(A11) $$r'' = \sqrt{(z_c - z_t)^2 + y^2}$$

(A12) $$f'' = \sqrt{f^2 + (j_t \cdot t)^2}$$

j_t = pixel row in image of top line

i_t = pixel column in image of top line

Formulae (A13) through (A15) can be written by inspection from Figure 5.

(A13) $$\tan \gamma_p = \frac{y}{z_c - z}$$

(A14) $$\frac{x}{i \cdot s} = \frac{r'}{f'}$$

(A15) $$\frac{y}{r'} = \sin \gamma_p$$

From (A14) and (A15) we obtain

(A16) $$x = i_p \cdot y \left(\frac{s}{r' \sin \gamma_p} \right)$$

which is formula (7) in the text.

From (A13) we obtain

(A17) $$z = z_c - y \cot \gamma_p$$

which is formula (6) in the text.

In formulae (A13) through (A17)

(A18) $$\gamma_p = 90^\circ - A_{12} - \phi$$

(A19) $$A_{12} = - \tan^{-1} \left(\frac{j_p \cdot t}{f} \right)$$

(A20) $$f' = \sqrt{f^2 + \left(j_p \cdot t \right)^2}$$

i_p = pixel column in point source flood image

j_p = pixel row in point source flood image

Aknowledgements

The authors wish to thank Marty Auman for assistance in preparing the manuscript.

References

1. Agin, G. J.," Real time control of a robot with a mobile camera." Technical note 179, SRI International, February 1979.

2. Bolles, R. C. and M. A. Fischler, " A Ransac-based approach to model fitting and its application to finding cylinders in range data. Proc. Seventh IJCAI, August 1981, 637-633.

3. Nagel, R. N., G. J. VanderBrug, J. S. Albus, and E. Lowenfeld, "Experiments in part acquisition using robot vision." Proc. Autofact II, Robots IV Conference, Detroit, MI. October 1979.

4. Nevatia, R. and T. O. Binford, "Description and recognition of curved objects. Artificial Intelligence 8 1. 1977, 77-98.

5. Oshima, M. and Y. Shirai, "Object recognition using three-dimensional information. Proc. Seventh IJCAI, August 1981, 601-606.

6. Popplestone, R. J., C. M. Brown, A. P. Ambler, and G. F. Crawford," Forming models of plane- and cylinder-faceted bodies from light stripes." Proc. Fourth IJCAI, September 1975, 664-668.

7. Posdamer, J. L. and M. D. Altschuler," Surface measurement by space-encoded projected beam systems." Computer Graphics and Image Processing 181, 1982, 1-17.

8. Rosenfeld, A. R. and A Kak, Digital Picture Processing, Academic Press, New York, 1976.

9. Sugihara, K., " Range-data analysis guided by a junction dictionary." Artificial Intelligence 12, 1, 1979, 41-69.

10. Mansbach, P., Calibration of a camera and light sources by fitting to a physical model. (In preparation).

j_b	$z_c \tan \gamma_b$	$\frac{s.r'}{f'}$
-122	5.3	.0702
-121	5.4	.0705
-120	5.5	.0707
.	.	.
.	.	.
.	.	.
0	21.3	.1169
.	.	.
.	.	.
.	.	.
+122	100.1	.348

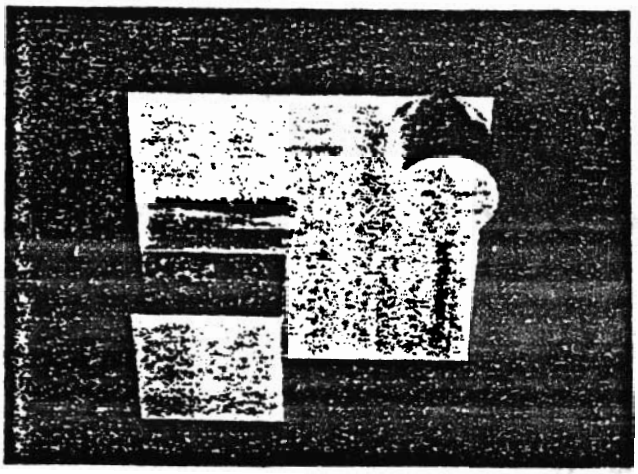
Table 1. Used to compute x and y coordinates of points in image of bottom line.

j_t	$\left(z_c - z_t\right) \tan \gamma_t$	$\frac{s.r''}{f''}$
-122	4.7	.0632
-121	4.8	.0634
-120	4.9	.0636
.	.	.
.	.	.
.	.	.
0	19.2	.105
.	.	.
.	.	.
.	.	.
+122	90.41	.313

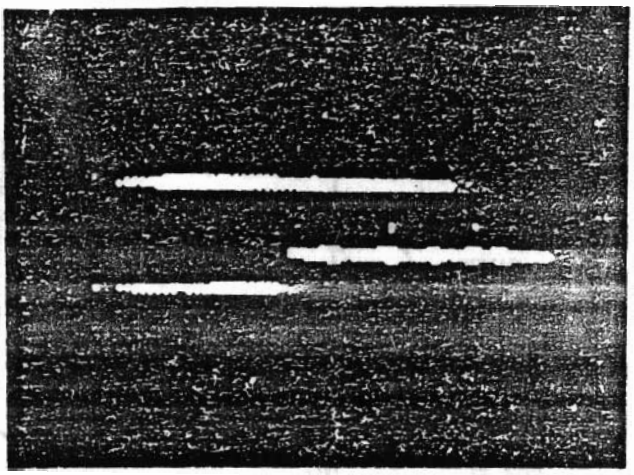
Table 2. Used to compute x and y coordinates of points in image of top line.

j_p	$\cot \gamma_p$	$\frac{s}{f' \sin \gamma_p}$
-122	3.84	.0135
-121	3.71	.0131
-120	3.66	.0129
.	.	.
.	.	.
.	.	.
0	.939	.00549
.	.	.
.	.	.
.	.	.
+122	.200	.00346

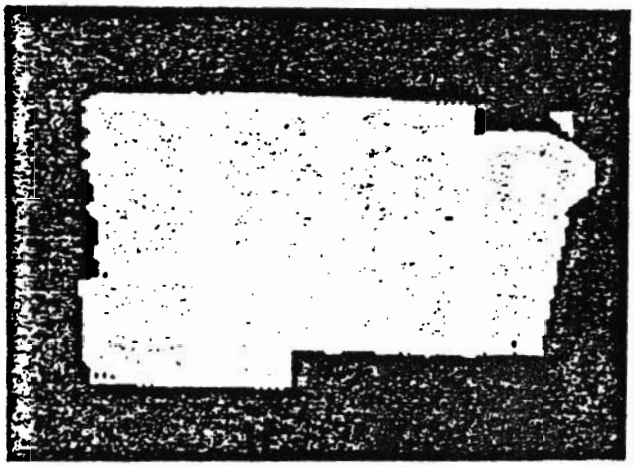
Table 3. Used to compute coefficients in equation (6) and (7) for point source flood image.



Photograph 1. A shiny metal object which is typical of parts in a machine environment. The image appears to be upside down due to the overhead position of the camera and the fact that the light sources are located beneath the camera. The position of the camera and lights relative to the object is illustrated in Figure 3.



Photograph 2. Image produced by the two planes of light. The break in the lower line is caused by the depth discontinuity of the raised square in the lower left of the object.



Photograph 3. Thresholded image produced by the point source flood light.

Photograph 4. Superimposition of image 2 and image 3.

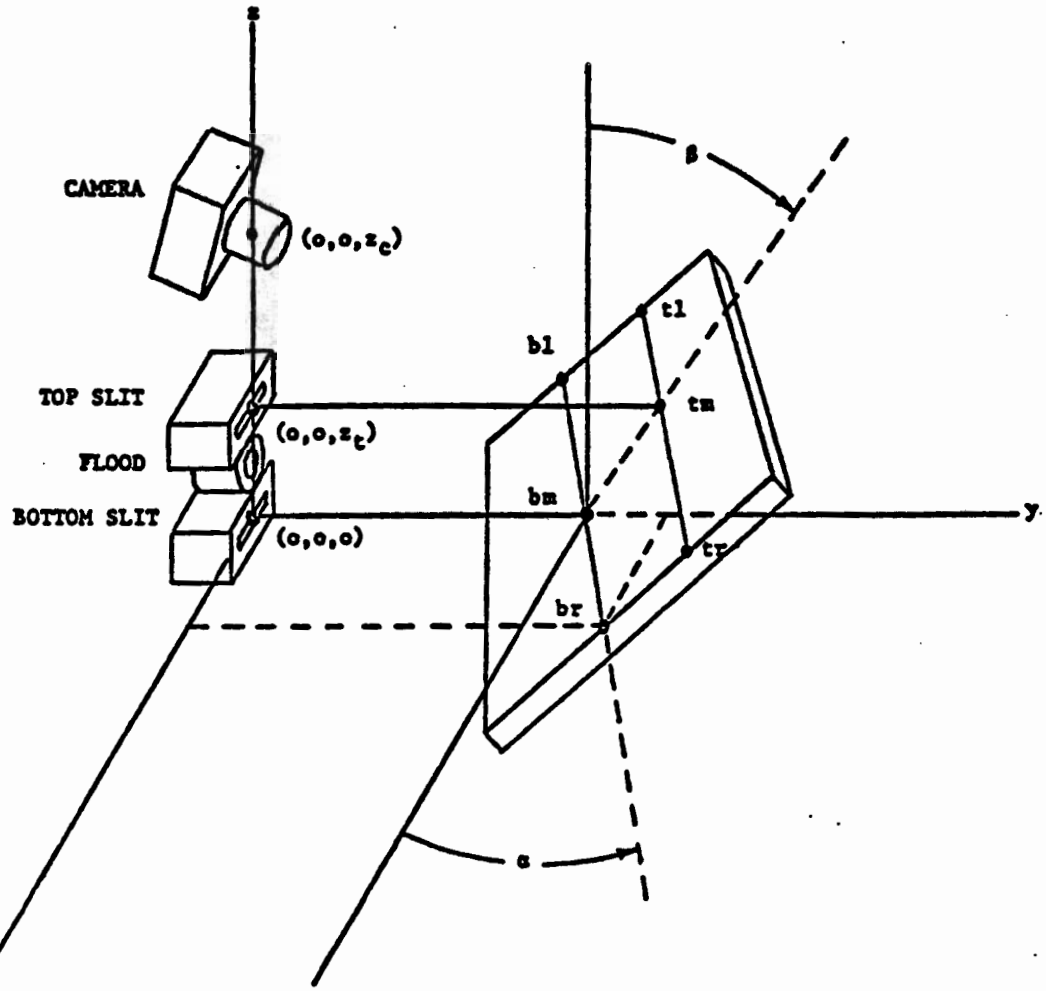


Figure 1. Relationsbips between the camera, the structured light sources, and a viewed planar object in a coordinate system fixed in the hand of the robot.

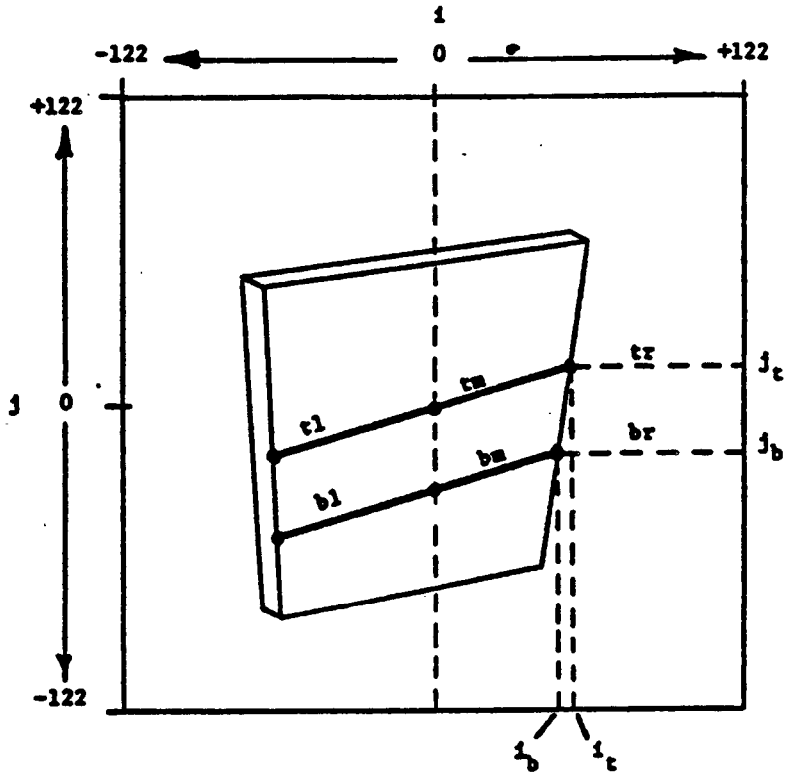


Figure 2. View of the object as seen by the camera illuminated by the pair of planes of light.

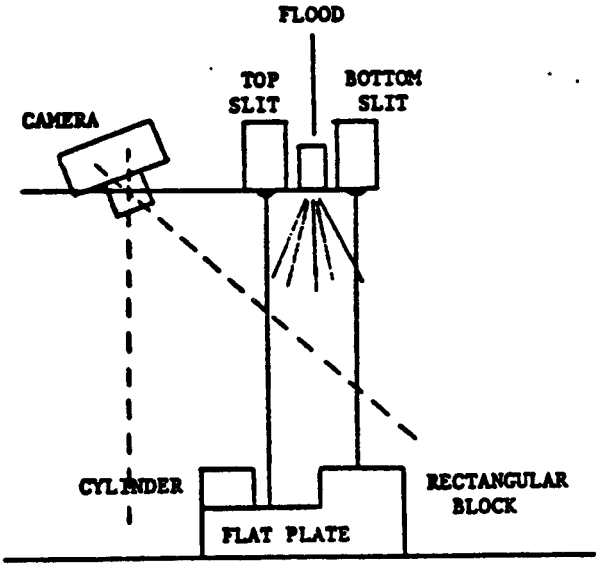


Figure 3. Viewing position of camera and light sources for photographs 2-5.

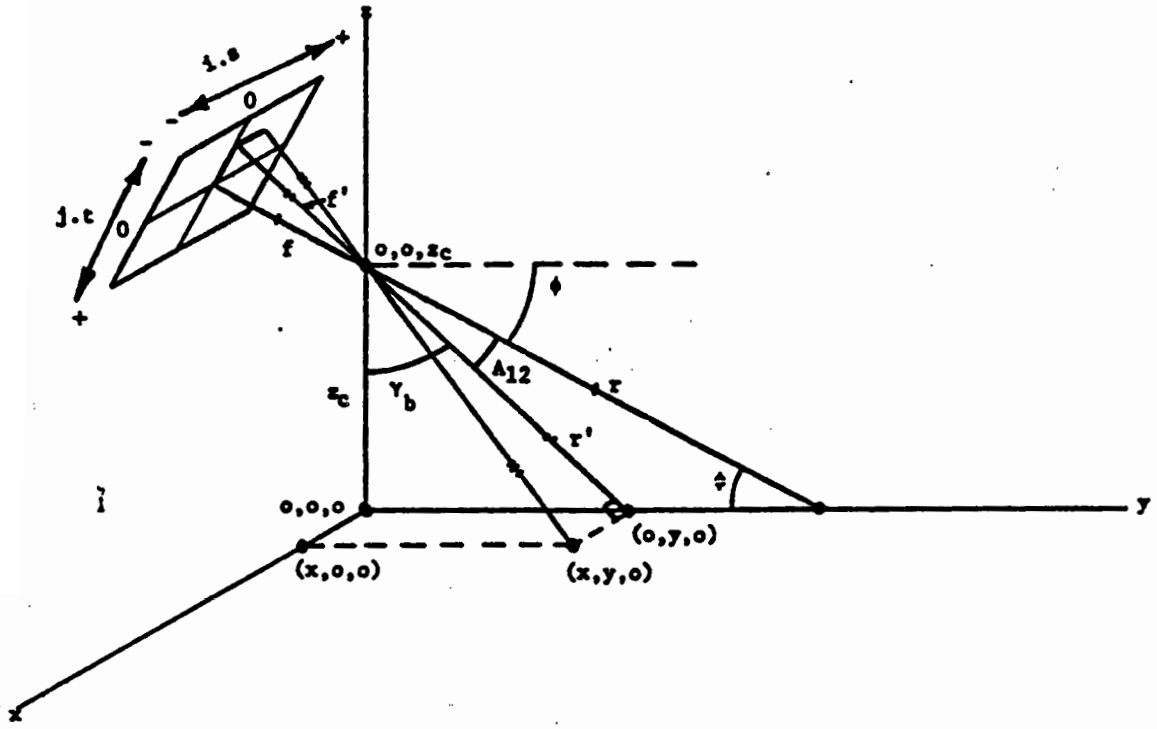


Figure 4. Relationships of camera pixels relative to illuminated points in the bottom plane of light.

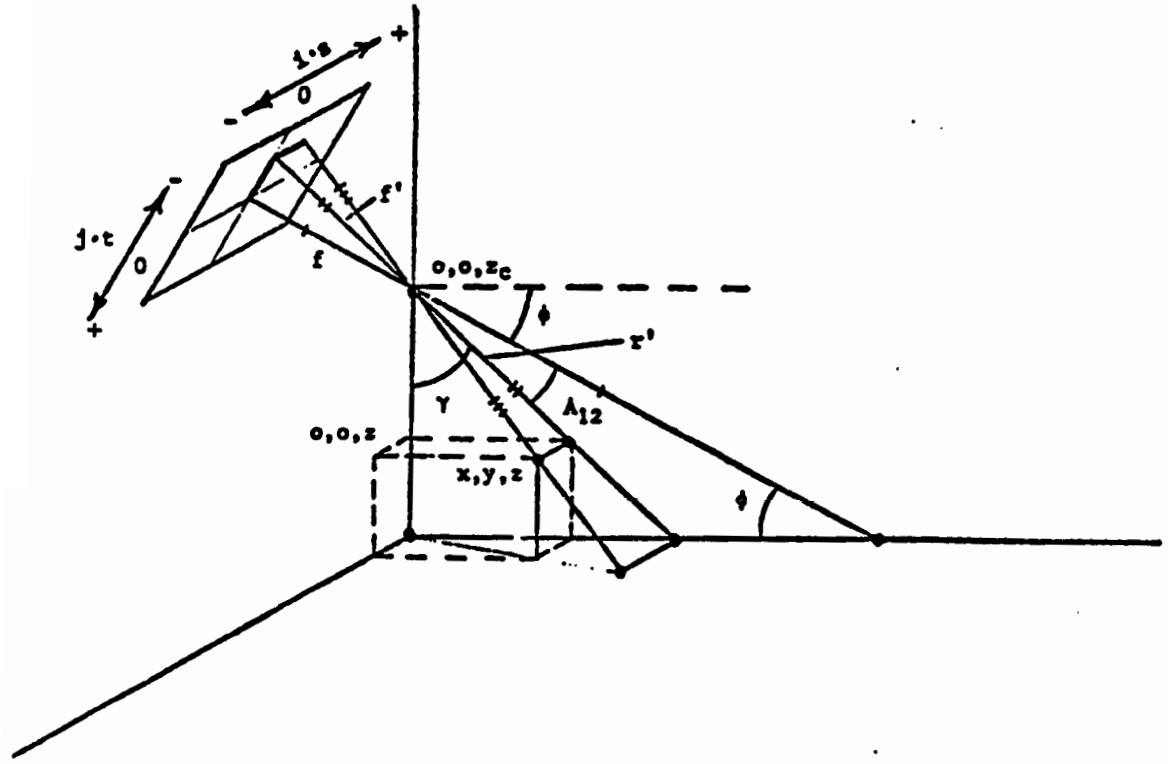


Figure 5. Relationships of camera pixels relative to arbitrary points on an object illuminated by the point source flood light.