MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

Recognition and Relocation of 3-Dimensional Objects Using Stereo Robot Vision System

Kyoung Mihn Do, Chee Woo Kang, and Kwae Hi Lee
Dept. of Elec. Eng., Sogang University

C. P. O. Box 1142
Seoul, Korea

ABSTRACT

An important task in stereo robot vision is to determine the three-dimensional locations of objects. The real location of an object is determined from the disparity of matched points in stereo image pairs.

This paper depicts a new stereo correspondence technique to match planes extracted from stereo image pairs. The stereo image pairs are taken from the lateral stereo camera model of eye-on-hand configuration with a camera on the end-effector of the robot.

Stereo Robot Vision System (SRVS) which finds out the three-dimensional locations of objects is described. Its software architecture consists of five major steps: stereo camera modelling, feature extraction, stereo image matching, depth determination, and interpolation.

In order to realize a knowledge-based robot which has the capabilities of cognition and problem solving, we make the robot recognize objects and relocate them in the specified order using the three-dimensional locations obtained from the SRVS. We show results of the SRVS with the proposed matching algorithm and the relocation work of the robot.

INTRODUCTION

The stereo image analysis in robot applications generally consists of following steps: stereo image acquisition, camera modelling, feature extraction, image matching, depth determination, and interpolation. Each step is important for determining accurate three-dimensional locations of objects.

There are several stereo camera models for taking the stereo images from the multi points of view known as the lateral stereo camera model, the axial stereo camera model and so on. We took the stereo image pairs from the lateral stereo camera model to make the three-dimensional robot vision system more close to human visual system and to use epipolar constraint. That is realized through having the robot, which has a camera on its end-effector, move along the planned lateral path.

Edges, segments, and zero-crossing points are often used as features for image matching. Many matching algorithms have been developed which were based either on the intensity-based stereo matching technique or on the feature-based stereo matching technique [1-3]. For we focused on the robot stereo vision for objects which are simple, not transparent, and time invariant, we proposed a new feature-based matching algorithm which matches planes and applied it to the stereo robot vision system (SRVS). The SRVS is a software system which determines three-dimensional locations of objects through several steps: stereo camera modelling, feature extraction, image matching, depth determination, and interpolation. The robot recognizes objects using the stereo image pairs, guesses the sequence of relocation, and relocates them in the order of height using the three-dimensional positions of objects extracted by the SRVS.

In this paper stereo image acquisition by eye-on-hand camera configuration, the SRVS, recognition and relocation, implementation and results, and conclusions are described.

IMAGE ACQUISITION AND CAMERA CONFIGURATION

In order to take the stereo image pairs from the lateral stereo camera model, the robot makes a planned lateral movement without any rotation and orientation having a camera on its end-effector which is far from the base. At the end points of lateral path, left image and right image are grabbed.

To determine the three-dimensional location from the stereo images taken from eye-on-hand configuration, it is necessary to know two relations between the camera and the robot base and between the object and the camera as shown in Fig. 1 [4]. The relation between the object and the camera known as the camera modelling is dealt in the next section. Notice in Fig. 1 that the transformation from the camera to the robot world is equivalent to rotation γ, β, and α degrees around the z, y, and x axes respectively, followed by translation from the origin of the camera frame to the robot world. Thus the homogeneous transformation matrix T_R^C from camera to the robot world coordinate system is given by

$$T_R^C = Trans(X_{co}, Y_{co}, Z_{co}) Rot(x, \alpha) Rot(y, \beta) Rot(z, \gamma)$$
$$= \begin{bmatrix} N_x & O_x & A_x & X_{co} \\ N_y & O_y & A_y & Y_{co} \\ N_z & O_z & A_z & Z_{co} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

,where (X_{co}, Y_{co}, Z_{co}) is the origin coordinate of the camera frame with respect to the robot world coordinate system.

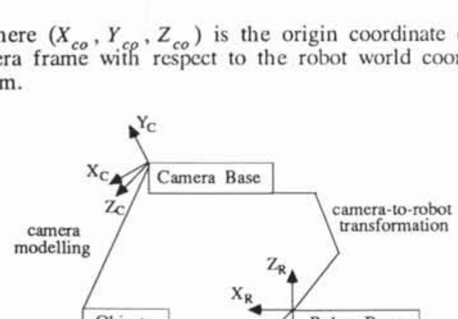


Fig. 1 Geometric relations in robot work space.

STEREO ROBOT VISION SYSTEM

Camera Modelling: In the lateral stereo camera model stereo images are taken from the two cameras which are separated only a translation b (length of the base line) in the x-direction without any rotation and orientation (see Fig. 2). Let the projection of a point $P(x_P, y_P, z_P)$ in three-dimensional space onto the left image plane and onto the right image plane be $P_L(x_L, y_L)$ and $P_R(x_R, y_R)$ respectively. From the geometric relations shown in Fig. 2 two relations are obtained [5]. All point coordinates are expressed in the left image coordinate system (x, y, z).

$$\frac{x_P}{x_L} = \frac{y_P}{y_L} = \frac{f - z_P}{f} \tag{2}$$

$$\frac{x_P + b}{x_R + b} = \frac{y_P}{y_R} = \frac{f - z_P}{f} \tag{3}$$

,where f is focal length.

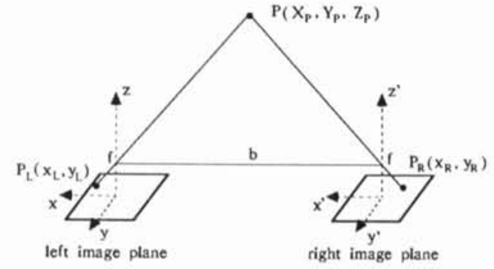


Fig. 2 The lateral stereo camera model.

From the epipolar constraint, the y coordinates of the two corresponding image points are equal, that is, $y_L = y_R$. Thus the point $P(x_P, y_P, z_P)$ in three-dimensional space is represented in terms of $P_L(x_L, y_L)$ and $P_R(x_R, y_R)$ as follows:

$$x_P = \frac{bx_L}{x_R - x_L} \tag{4}$$

$$y_P = \frac{by_L}{x_R - x_L} \tag{5}$$

$$z_P = f\frac{x_R - x_L - b}{x_R - x_L} \tag{6}$$

Combining equations (1) and (4), (5), and (6) to obtain the three-dimensional location of object relative to the robot base using the disparity of the matched points, we obtain

$$x_{PR} = \frac{1}{x_R - x_L}[b(N_x x_L + O_x y_L) + fA_x(x_R - x_L - b)] + X_{co} \tag{7}$$

$$y_{PR} = \frac{1}{x_R - x_L}[b(N_y x_L + O_y y_L) + fA_y(x_R - x_L - b)] + Y_{co} \tag{8}$$

$$z_{PR} = \frac{1}{x_R - x_L}[b(N_z x_L + O_z y_L) + fA_z(x_R - x_L - b)] + Z_{co} \tag{9}$$

,where (x_{PR}, y_{PR}, z_{PR}) is the position of the point P in three-dimensional space with respect to the robot world coordinate system.

Plane Extraction: Planes are extracted through the following four steps: edge extraction, filtering, thinning, and tracking. The technique developed by Burns, Hanson, and Riseman is used in this paper to extract edges from images [6]. If any peak or hole is near an edge, the edge becomes complex, ambiguous, and distorted after thinning. So we suggest a filtering technique which cuts peaks and fills holes near the edges. After filtering, more clear and less distorted edges were obtained. Then planes are extracted by region growing technique and six parameters are found out.

The center, area, and perimeter of a plane are stored as parameters. The short (long) cut is defined as the minimum (maximum) distance from the center to boundary of the plane. Brightness is the average intensity of pixels in the plane. These six parameters are used as important features to match planes.

Plane Matching: Many different feature-based image matching techniques has been developed. Features which generally used in those techniques are linear features like lines [6]. The technique developed by J. H. McIntosh and K. M. Mutch matches straight lines well [2]. But this technique has three major disadvantages for applying to robot vision. The first problem is distortion of image. The second is the inaccuracy of the computed disparity of lines that are displaced in a direction similar to the angle of the lines. And the last is the considerable computation time in calculating eight parameters of lines.

Considering the above disadvantages and focusing on stereo robot vision for simple objects, we suggest a new feature-based matching technique based on the straight line matching technique. The new matching technique in this study finds corresponding pairs of planes from the stereo images based on difference values, a match function, and thresholds. The only plane which is matched mutually is considered to be matched. The three thresholds, the maximum row distance, the maximum column distance, and the match threshold, are required for matching planes. The match threshold is used to control the minimum acceptable value of a match function and 0.7 is normally used for the match threshold value. The maximum row distance and the maximum column distance are used to determine the possibility of two planes to be matched and these values are determined from the epipolar constraint.

The difference value is defined as the difference of the maximum of parameter values and the minimum of parameter values for the two planes. The difference value $DIFF_i(L, R)$ is defined as

$$DIFF_i(L, R) = max[Para_i(Plane_L), Para_i(Plane_R)] - min[Para_i(Plane_L), Para_i(Plane_R)]$$

$$i = \text{area, short cut, long cut, perimeter}$$
$$L = 1, 2, \cdots, N$$
$$R = 1, 2, \cdots, M$$

,where N is the number of planes in the left image and M is the number of planes in the right image.

A difference threshold is used for selecting a $Plane_L$ or a $Plane_R$ as a candidate for plane to be matched corresponding to a $Plane_R$ or a $Plane_L$ and defined as

$$DIFF\_TH_i(L, R) = \min[Para(Plane_L), Para(Plane_R)]w_i$$

,where w_i is the weight of the ith parameter.

The weight represents the relative importance of a plane and has a value greater or equal to 0 and less or equal to 1. Choosing the best weights for planes is a matter of trial and error and the weight value normally used is 0.3.

The match function only for the brightness parameter is defined as the ratio of the minimum and the maximum brightness parameter value in the two planes, that is

$$MF_{LR} = \frac{min\,[brightness\,(Plane_L),\; brightness\,(Plane_R)]}{max\,[brightness\,(Plane_L),\; brightness\,(Plane_R)]}$$

Two arrays which holds the best match of each plane from each image are defined as

$$MAX_L = max\,(MF_{Lj}) \qquad j = 1, 2, \cdots, M$$
$$MAX_R = max\,(MF_{iR}) \qquad i = 1, 2, \cdots, N$$

The best matches are determined as follows:

IF the differences of x coordinate and y coordinate between the center of a $Plane_L$ and the center of a $Plane_R$ are less than the maximum row distance and the maximum column distance,
AND IF $DIFF_i(L, R) < DIFF\_TH_i(L, R)$,
AND IF $MF_{LR} = MAX_L$ AND $MF_{LR} = MAX_R$,
AND IF $MF_{LR} >$ match threshold,
THEN Match (L, R).

Depth Determination and Interpolation: The three-dimensional location of an object with respect to the robot world base is determined by using the result of image matching and equations (7), (8), and (9). A pixel of the matched plane is matched to a pixel in a corresponding matched plane by linear interpolation. Then by deriving the x, y coordinates of two matched image points in the stereo images to equations (7), (8), and (9), we obtain the three-dimensional location of the object. The pixels out of planes are not considered to be matched by interpolation but regarded as working environment.

RECOGNITION AND RELOCATION

Recognizing objects is necessary before relocating the objects in three-dimensional space. The first step in recognition is to search planes to know what planes are belong to an object and the next step is to recognize the object from the relation between planes. Three-dimensional informations of objects which are the position (x, y coordinate with respect to the robot world coordinate system) and the height (z coordinate) in three-dimensional space are obtained using the three-dimensional locations of planes determined from the SRVS.

The relocation work of the robot in this study which is relocating objects in the order of height is not a simple work such as moving objects to the prescribed positions but a knowledge-based work such that the highest object is relocated at the position which has the maximum coordinate in the y direction relative to the robot world base at which another object had placed. The robot must guess the relocation sequence according to the configuration of objects to do the relocation work with minimum movements.

IMPLEMENTATION AND RESULTS

We implemented the knowledge-based robot system with stereo vision system using a industrial FUNAC Arc Mate robot, IBM PC/AT, and a standard resolution camera. It took 8 minutes from obtaining the stereo image pairs of objects to relocating objects. The knowledge-based robot system in this paper is fully automatically controlled by the computer. The unique manual work is to mount a camera at the end-effector of the robot. Fig. 3 shows the knowledge-based robot system which is realized in this study.

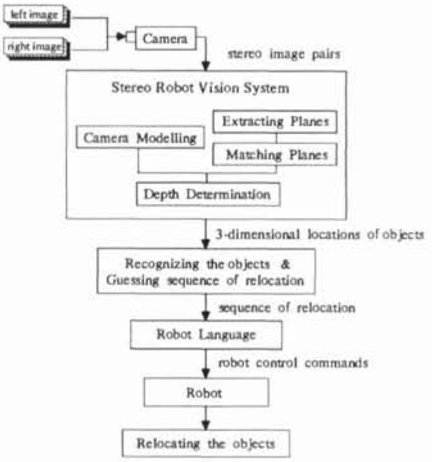
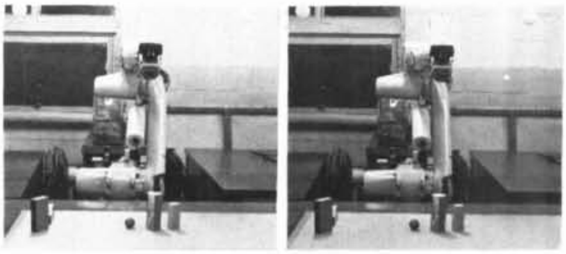


Fig. 3 The knowledge-based robot system implemented in this paper.

The robot moved along the planned lateral path which was 3 *cm* long. And as shown in Fig. 4 the stereo images were taken at the end points of the path.

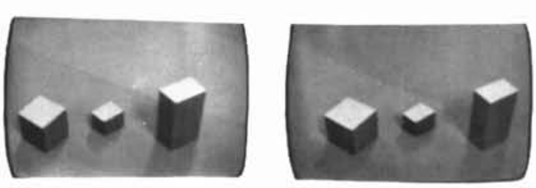


left image acquisition right image acquisition

Fig. 4 Eye-on-hand camera configuration and stereo image acquisition from the lateral movement.

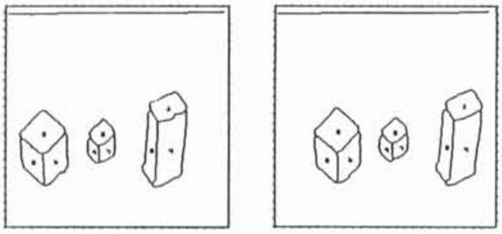
The new image matching technique we proposed has a problem that determination of weights of parameters and the match threshold is difficult, however it matched all planes in the stereo images correctly and had fast processing time of average 42 seconds which is less than the processing time (171 seconds) of the straight line matching technique. So this technique is efficient to use in stereo robot vision for simple objects. Fig. 5 shows one of the input stereo image pair taken from the lateral stereo camera model. The images in Fig. 6 are the results of the new image matching technique. The number indicates pairs of matched planes in the left image and the right image. Fig. 7 shows the three-dimensional locations of three objects which was determined from the SRVS.

The average error in the three-dimensional locations of objects obtained from the SRVS is 0.89 *cm* while the maximum error is 2.52 *cm*. There are three causes of the location extraction error. The first one is that the camera was not located accurately at the end-effector. So



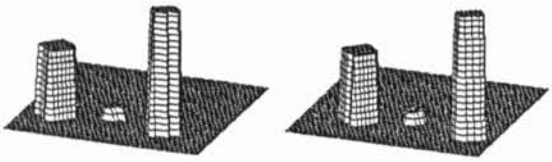
left image right image

Fig. 5 A stereo image pair.



left image right image

Fig. 6 The result of the plane matching technique.

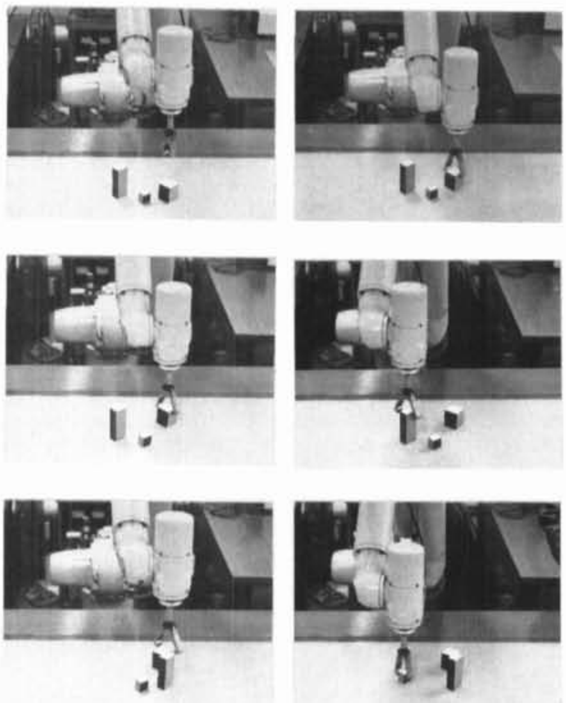


left image right image

Fig. 7 Three-dimensional locations of three objects, the output of the SRVS.

the camera modelling was errorneous for geometric relation between the camera and the object. It is another cause that the calibration of focal length of the camera and the origin of the camera coordinate with respect to the robot world coordinate system were not accurate. And the last cause is distortion in epipolar line and in boundaries (edges) of planes.

The relocation sequence was determined correctly according to the configuration of objects by the robot. Although the robot did not grip the center of object due to the errors in the three-dimensional locations obtained from the SRVS, the robot recognized objects correctly and relocated them in the order of height with minimum movements successfully. Fig. 8 shows that the robot is relocating objects in the order of height.



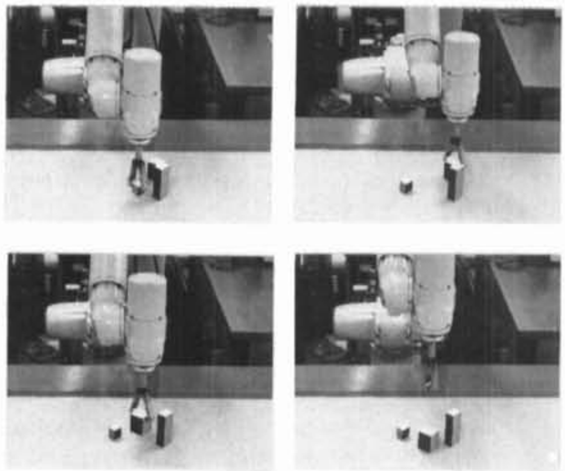


Fig. 8 The robot relocating objects in the order of height with minimum movements.

CONCLUSIONS

We realized a knowledge-based robot system containing a stereo robot vision system which determines the positions of objects in three-dimensional space, recognizes objects, and relocates them in the specified order by guessing the sequence of relocation. And we proposed a new feature-based image matching technique which has the merits of speed, accuracy, and efficiency to use in stereo robot vision for simple objects.

To find out more accurate three-dimensional locations from the SRVS, the error analysis for camera modelling and calibration is required. The knowledge-based robot system will be useful for works like position corrections, undesirable element changes, and other applications in robotics.

REFERENCES

[1] M. A. Gennert, *Brightness-based stereo matching*, Second International Conference on Computer Vision, pp. 139-143, 1988

[2] G. Medioni and R. Nevatia, *Matching images using linear features*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 6, No. 8, pp. 676-685, Nov. 1984.

[3] J. H. McIntosh, *Matching straight lines*, Computer Vision, Graphics, and Image Processing, vol. 43, pp. 386-408, 1988.

[4] R. K. Lenz and R. Y. Tsai, *Calibrating a cartesian robot with eye-on-hand configuration independent of eye-to-hand relationship*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 11, No. 9, pp. 916-928, Sep. 1989.

[5] N. Alvertos, D. Brazakovi, and R. C. Gonzalez, *Camera geometries for image matching in 3-D machine vision*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 11, No. 9, pp. 897-915, Sep. 1989.

[6] J. B. Burns, A. R. Hanson, and E. M. Riseman, *Extracting straight lines*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 8, No. 4, pp. 425-455, July 1986.