PRECISE MATCHING OF MOTION STEREO IMAGES GUIDED BY COARSE DISPARITY MAP

Gang Xu, Saburo Tsuji and Minoru Asada
Department of Control Engineering
Osaka University

ABSTRACT

A stereo matching method is proposed which utilizes a new coarse-to-fine rule to overcome the difficulties resulted from occlusion. At first, a pair of stereo images are taken for a short base line distance for which occulusion seldom occurs. A matching is then assigned to the pair and the result (a disparity map) is used to make a plan to obtain more accurate distance map of the scene; the system takes stereo images for longer camera distances and finds correspondences of edges in each pair of images guided by the disparity map obtained already. The matching process is efficient and reliable because the locations of edges and possibilities of occulusion are predicted.

1. INTRODUCTION

Early researches on stereopsis [1] have pointed out that the problem of false targets lies at the heart of the matching problem, and both a wide disparity range and a fine resolution enhance the severity, and the need for a sophisticated method that acquires both range and resolution has been clarified.

Marr and Poggio have presented a zero-crossing algorithm[2] (implemented by Grimson [3]) which is based on the so-called coarse-to-fine rule. The stereo images are filtered at different scales to yield a set of primitive image pairs of different resolutions. A match of a wide disparity range but of a coarse resolution is obtained at first and the result is then used to restrict search ranges for a finer resolution.

Their ideas provide us with a very powerful tool for obtaining 3-D information in scene, however difficulties in matching still exist if we wish to acquire accurate range data from stereopsis. A large camera distance is needed for the accuracy, but many parts of one image are invisible in the other image by occlusion for such arrangement of cameras. Thus, the matching is very difficult and unreliable. Sophisticated edge-based stereo algorithms have been studied to solve the problem [4].

This paper presents a new idea to overcome the above-mentioned difficulties. We use a motion stereo (slide stereo) and a new coarse-to-fine rule is employed. At first, a pair of stereo images are taken for a short base line distance for which the occlusion seldom occurs. A matching is assigned to the pair, then the result (disparity map) is used to plan how a more accurate distance map will be obtained. A set of images are taken for larger camera distances and the matching is established between edges in each pair of images guided by the disparity map obtained already. The matching process is efficient and reliable because the locations of edges and possibilities of occlusion are predicted.

2. PLANNING OF CAMERA MOVEMENT

It is well known that the accuracy in measuring the range from stereo images is proportional to the base line distance. A slide stereo (a pair of stereo images are taken by moving a camera along a line) will obtain the best estimate of the distance to a point in the scene for a camera arrangement that the disparity of the point is maximum, the point appears very close to the boundary of the both images. We propose that the system plans both the camera movements and image analysis based on the results obtained already.

The system iterates the process of move, imaging and analysis. After taking the first image, the camera is moved by a predetermined short distance and take the second images. The first step of matching is assigned to these stereo image, and we obtain a disparity map of a small disparity range but of a fine resolution, where the edges are detected by a Laplacian-Gaussian filter of which the center region width W is 4. Since the camera distance is small, occulusion seldom occurs in the image pair.

Now the system makes a plan to move the camera to obtain a more accurate depth map. In order to simplify the matching process, we limit the length of camera movement at each step.

Suppose that the right camera is moved at the i+1-th step. Let the base line distance of step i+1 be k times that of step i. Then if the disparity of an edge point at $(X_r^{\,i},Y)$ in the right image is D^i, the location $(X_r^{\,i+1},Y)$ of the edge point at the i+1-th step is determined as

$$X_r^{\,i+1}=X_l^{\,i+1}-kD^i \qquad (1).$$

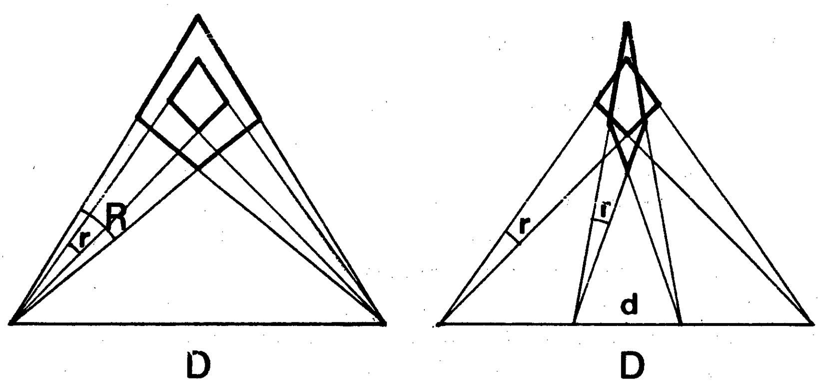
Taking into account the error in D^i, eq(1) should be revised:

$$X_r^{\,i+1}=X_l^{\,i+1}-kD^i+k\Delta D \qquad -|\Delta D|_{max}\leq\Delta D\leq|\Delta D|_{max} \qquad (2)$$

where $|\Delta D|_{max}$ is usually not greater than 1 pixel. Since the Gaussian-Laplacian filter of w=4 is used, we search a region with the size is 5 pixels for the matching process. Therefore, if k=2 is adopted, reliable matching is expected.

3. NEW COARSE-TO-FINE RULE

Both our algorithm and that of Marr and Poggio are based on the coarse-to-fine rule, of which the target is to obtain both a wide disparity range and a fine resolution. Their routes to attain the goal, however, are different. Fig.1 helps to understand the difference. The feature of the method by Marr and Poggio is that the accuracy in depth improves with the resolution getting finer, while the disparity range remaining unchanged. The accuracy in depth of our method improves with the disparity range getting wider, while the resolution remaining unchanged.



(a) (b)

Fig.1 Two routes to improve the depth accuracy

One important advantage of our method over Marr and Poggio's is that the possibility of mismatch resulted from occlusion is very small. Another advantage is the system selects the best camera arrangement based on the disparity map obtained at the earlier stage. The demerit of the proposed method is to take longer time for iterating camera movement and stereo matching.

4. OCCLUSION PREDICTION

As described in Section 3, the edge location at the i+1th step is predicted from the disparity map of the ith stage. Now let us consider how we can predict the occlusion.

Suppose there are two edge points P_1 and P_2 on a horizontal scan line in the right image at the i-th step, and the right camera is moved at the i+1-th step. If the points are predicted as changing their order to appear on the horizontal line in the i+1-th image, or

$X_r{}^i{}_1 < X_r{}^i{}_2$,
$X_r{}^{i+1}{}_1 > X_r{}^{i+1}{}_2$,

then the possibility that P_1 will be occluded is high.

The prediction is not always true because of the error in estimates of edge positions. Therefore, the following strategy is adopted. If the distance between the occluding and occluded edges is larger than a threshold, then the occlusion is certain and we search the i+1th image only for the occluding edge (P_2 in this case) to match the edge pair. If the distance is smaller than the threshold, we search the region of the specified width for two edge points. If there exist two edge points, the left one corresponds to P_1 and the other is P_2. If only one edge point exists, then we consider that it corresponds to the nearer one P_2.

5. EXPERIMENTAL RESULTS

We have tested the described technique for images synthesized by computer and images of blocksworld. Figs.2 show the input images taken at five camera positions, (a), (b), .., (e) are pictures taken at the first, second,..., and the fifth positions, respectively. We use (a) as the left image and (b)-(e) are matched to (a) as the right images. Figs.3 show the zero crossing contours of Figs.2(a) and (e), the contours are displayed as bright if the gradient is positive and as medium grey for the negative gradient. Note that there is change of sign if a more distant object is occluded. Since the sign is used as a primitive of matching criteria, the change may cause erroneous results. Figs.4 show the disparity maps obtained by matching Fig.2(a) to (b) and (a) to (e). The lengths of horizontal bars represent the magnitudes of the disparities. Our experience assures that most edge points appear in the predicted search ranges, if the camera movement is accurate.

6. CONCLUSIONS

We have described a new coarse-to-fine rule for obtaining a precise depth map. The method has the advantage over the conventional methods that it is free from the mismatch resulted from occlusion.

There are many future problems for our method. The Laplacian-Gaussian filter may not be the best edge detection method for the images in which occlusion occurs. The adopted strategy of camera displacement (k=2) should be studied further. Another important future research theme is that the system autonomously makes more sophisticated plans for obtaining better input images for stereopsis. We are now investigating how a robot can plan the camera movement based on the disparity map.

REFERENCES

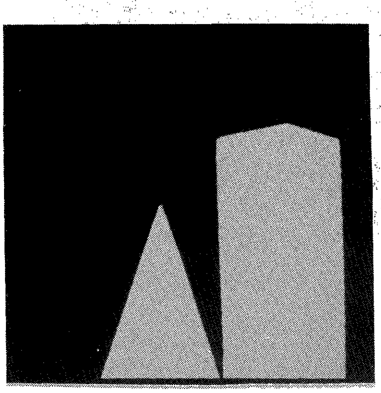
1. Marr,D. and Poggio,T. 1976, Cooperative Computation of Stereo Disparity,Science,N.Y.194:283-287

2. Marr,D. and Poggio,T.1979, A Computational Theory of Human Stereo Vision, Proc.R.Soc.Lond.B204:301-328

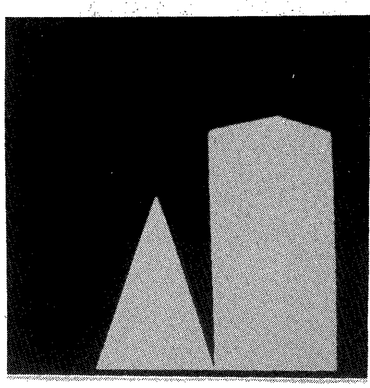
3. Grimson,W.E.L. 1981,From Images To Surfaces: A Computational Study of the Human Early Visual System,Cambridge,MA,MIT Press

4. Ohta,Y. and Kanade,T. 1983, Stereo by Intra- and Inter-scanline Search Using Dynamic Programming, Tech. Report,CMU-CS-83-162,Computer Science Dept.,CMU

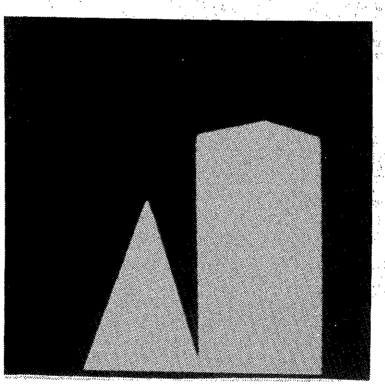
5. Marr,D. and Hildreth,E.1980, Theory of Edge Detection,Proc.R.Soc.Lond. B207:187-217



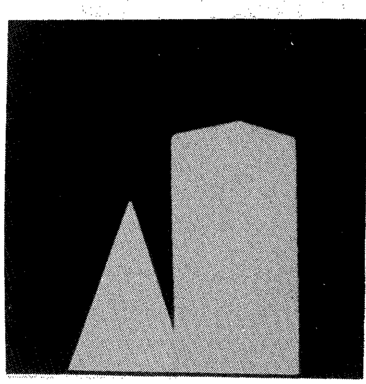
(a)



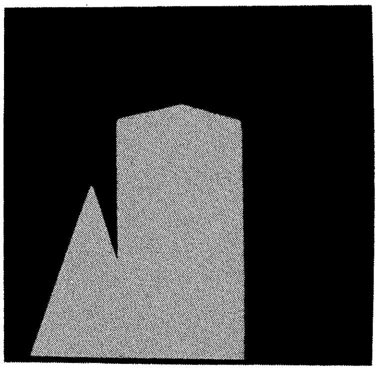
(b)



(c)



(d)



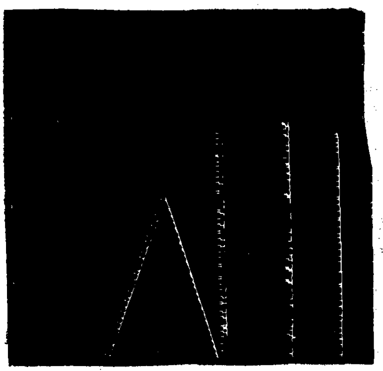
(e)

Figs.2 Input images

(a)

(b)

Figs.3 Zero crossing contours



(a)

(b)

Figs.4 Disparity maps