

MEASUREMENT OF DISPARITY IN STEREO
DEPTH PERCEPTION*

A. Goshtasby and C. V. Page
Dept. of Computer Science
Michigan State University

Measurement of disparity in stereo images of the same scene is approached by segmenting one image and selecting points on region boundaries for matching. The sources causing the mismatches are identified and counter-measures to avoid them are given. The proposed approach is tested against two sets of stereo images.

INTRODUCTION

Images obtained at different viewpoints from a three-dimensional scene will have geometric differences. Image disparity is referred to the geometric difference between two images of the same scene. Image disparity is the key factor in determination of depths of points in the scene.

If the images are obtained by a well balanced stereo camera system then disparity can be obtained by the difference of position of corresponding points horizontally, in the two images. If the images are obtained by one camera with displacement, then there is a need to align the two images for disparity measurement. Image alignment is possible if a number of points in infinity can be identified in both images. After points in the infinity are registered in the two images, the positional difference between other corresponding points in the images provide information about depths of points in the scene.

To be able to measure image disparity, we need to know the position of corresponding points in the images. Image point correspondence problem has been studied by different groups.

* This work has been supported in part by NASA Grant NGL-23-004-083 and NSF Grant MSE-08-04126.

Marr-Poggio-Grimson have tried to match zero-crossings that are oriented non-horizontally, to determine the position of corresponding points in the images [1,2]. Zero-crossings are defined as the second directional derivatives of intensities in the image and are used to locate intensity changes in an image.

Zero-crossings have been defined for four size masks in [1]. Zero-crossings from larger masks are used to establish global correspondence while zero-crossings from smaller masks are used to find local correspondence. Global correspondences have been used to resolve ambiguities among local correspondences.

Baker-Binford-Arnold have made attempts to match high gradient edges in the images for the purpose of image disparity measurement [3,4]. Since object boundaries usually involve intensity changes and intensity changes produce zero-crossings and high gradient edges, by matching zero-crossings or high gradient edges it is possible to measure depths of points on object boundaries in the scene and so be able to locate the objects.

Hannah-Moravec-Nevatia have approached the image point correspondence problem by window searching [5,6,7]. Window searching suffers from the fact that low variance areas of the image do not produce accurate matches. But in matching of high variance areas since a neighborhood is used in the search, the best match usually gives valid corresponding points in the two images. While in the matching of zero-crossings and high gradient edges, the best match doesn't necessarily give the position of truly corresponding points in the two image. The reason is that a scene viewed at two different viewpoints do not produce the same edges and the technique of zero-crossing or edge matching try to find the best match edges which may not necessarily be the corresponding ones.

In the following, a point correspondence technique is described which 1) selects points on region boundaries for matching and 2) carries out matching by a window search process. Region boundaries are highly informative areas of the image and measurement of disparity on region boundaries results in detection of depth discontinuities. Window search is employed as the searching strategy for sensible match. In the proposed technique, first one of the images is segmented and then an attempt is made to determine the position of corresponding points on region boundaries in the other image by the window search approach.

IMAGE SEGMENTATION

Image segmentation is the process of dividing an image into regions whose points have nearly the same property. This goal can be achieved in two ways. One way is by determining boundaries between homogeneous regions of different properties. Since the boundaries are places where there are sharp changes in property values, region boundaries can be found by an edge operator. Davis gives a review of image segmentation by this approach [8].

Another way to segment an image is by determining regions of homogeneous property. This can be done by thresholding the image at different levels and letting each region include those neighboring points which have property values between two given threshold values. A survey of image segmentation technique by thresholding can be found in [9]. Another way to obtain regions of homogeneous property in an image is by region growing. In this approach, neighboring homogeneous regions of similar property are joined, or nonhomogeneous regions are split into homogeneous ones to segment the images. A survey of image segmentation by region growing is given by Zucker [10].

We will segment an image by the thresholding approach. A threshold value is estimated using those gradient values which are both high in value and also numerous in the image. High gradient values are used for threshold estimation because high gradient values correspond to sharp changes in intensity values and this shows the boundary between two regions of different property. Since noise pixels also have high gradient values, to avoid this, the high gradient value which is numerous in the image is used.

If we obtain the gradient of an image and construct the histogram of the gradient image, we can determine the number of pixels in the image with a given gradient value. Let $G(j)$ show the number of pixels in the image with gradient j . Now if we define,

$$M(j) = j * G(j)$$

then those pixels with large j (high gradient) and large $G(j)$ (showing that a large number of them are available) will result in large value of $M(j)$. This tells us that we should be looking at high values of $M(j)$ to determine the threshold value. How about the maximum of $M(j)$?

Figure 1 shows three different cases of $M(j)$ depending on the contents of the image. Figure 1.a shows the case where the maximum value of $M(j)$ corresponds to the high gradient pixels in the image (j_1 is high), and determining the threshold value using gradient j_1 is proper. Figure 1.b shows the case where the

maximum of $M(j)$ does not correspond to high gradient edges (j_2 is not high). But $M(j)$ has another peak value at j_1 and j_1 is high and is proper for threshold estimation. Figure 1.c shows the case where $M(j)$ has one peak and it is at j_2 which is low and is not proper for threshold estimation. Using these facts, the following procedure is proposed for threshold estimation.

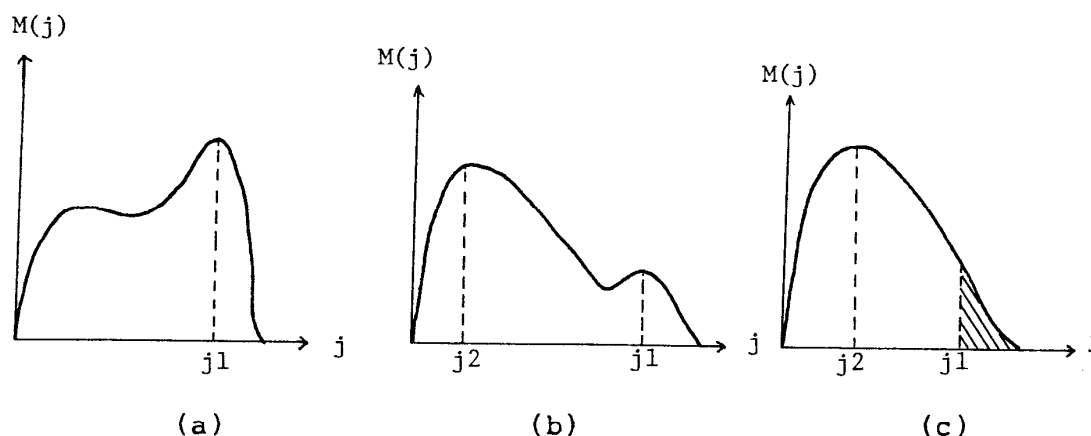


Figure 1. Three different histograms of $M(j)$.

1. Obtain the gradient of image 1 and construct its histogram, $G(j)$.
2. Compute $M(j)=j \cdot G(j)$ for all j , and smooth $M(j)$ to avoid noisy peaks (we have smoothed in 5 neighborhoods).
3. Determine the rightmost peak of $M(j)$. Let the peak be at $j=j_1$.
4. Compute $SUMG = \sum G(j)$. If $SUMG \leq 5\%$ of total image (total image = $\sum G(j)$), then
 Threshold value = average intensity of pixels with gradient j_1
 Otherwise (when $SUMG > 5\%$ of total image),
 Threshold value = average intensity of pixels at the 5% highest gradient line in the image.

Step 3 assumed that $M(j)$ has at least one peak value. $M(j)$ will in fact have at least one peak because it starts at zero (when $j=0$) and returns to zero (when $G(j)=0$). The constant 5 neighborhoods and 5% which were used in steps 2 and 4, respectively, were determined heuristically to best fit our set of imagery. For other images, this might need modification for best results.

When segmenting an image, it is possible that several objects at different depths in the scene be segmented into one region. This makes the determination of position of some objects impossible. In images where more than two types of regions are present, one threshold value is not able to segment the image into regions of

different types. For such cases, however, the above segmentation technique can be applied recursively to regions in the image that have variances above a threshold value.

DISPARITY ON REGION BOUNDARIES

To be able to determine the disparity of points belonging to region boundaries, we have to locate the position of points on region boundaries in the two images. If two balanced, equal focal length cameras are arranged with axes parallel, then it can be assumed that they share a single common image plane. Any point in the scene will be projected into two points on the joint image plane. If we connect these two points, the obtained line will be parallel to the cameras' baseline (baseline: a line connecting the cameras lens centers). This shows that given a point in one of the images, we can determine its corresponding point in the other image by searching along a narrow band passing through the given point and parallel to the baseline.

At this stage we use an image with region boundaries overlaid and an original image from the other camera for window searching. For each boundary line, first we have to establish a pair of corresponding points in the two images. To do so, we take a window centered at an arbitrary point on the given boundary and search it in the other image by shifting the window in a narrow band parallel to the baseline passing through the selected point. At each shift position, the similarity between the selected window and the window over which it is lying is determined. The shift position giving the largest cross-correlation value is taken as the true match and their centers are taken as corresponding points in the two images.

When one pair of corresponding points in the two images are determined, by following the boundary, we take windows centered at the boundary points and search them in a small neighborhood of the previous corresponding point to locate them. Since the domain of search is limited to a narrow band parallel to the baseline and to the allowable disparity difference of a point and its neighbor, we need to search only in small areas (like 3×7). This reduces the false match probability and also increases the speed of the disparity measurement process.

Since disparity difference of two neighboring points which belong to different objects at different depths in the scene could be larger than expected, on those occasions where a match rating smaller than a threshold value is obtained, we will increase the search domain and search the same window in a larger area (like 3×17) for security. In this manner, the position of points on region boundaries can be located in the other image. Disparity

on a region boundary is simply the horizontal difference between corresponding points on the boundary in the two images.

DISPARITY INSIDE REGIONS

Areas inside regions have low intensity variances and window searching will not be accurate. Even not so, since points are mostly inside regions than on region boundaries, carrying out window search for every point inside regions makes the image disparity measurement slow. However, we may be required to determine the disparity of points inside regions also. If the objects in the scene are known to have planar surfaces, then since disparity on a plane is well characterized, we will be able to determine disparity of points inside regions by interpolating disparity of points on their boundaries. If disparity of two points A and B on a plane are d_1 and d_2 , respectively, then the disparity of point C on line AB will be equal to,

$$d_2 + (d_1 - d_2)k / (h + k)$$

where h and k are distances of C to A and B, respectively (see Figure 2).



Figure 2. Determination of disparity inside regions.

If geometric models for objects in the scene are known a-priori, and if the objects can be easily recognized, then by fitting the right model to the points on an object boundary, we can estimate disparity of other points on the object also. But if no a-priori knowledge about objects in the scene are known, since homogeneous areas usually correspond to planar surfaces, approximation of disparity inside regions by linear interpolation would be acceptable in some applications.

AVOIDING MISMATCHES

To be able to avoid the mismatches, in the following, the sources that cause the mismatches are investigated first. Then attempts are made to suppress the cause for mismatches. The following

sources are identified as major cause for mismatches.

1. Occluded points.
2. Geometric differences.
3. Homogeneous areas.
4. Weakness of the similarity measure.

OCCLUDED POINTS

An occluded point is a point which can be seen from only one of the cameras. So, if a window centered at the occluded point is taken from one image and is searched in the other image, always a mismatch will be obtained. Because the point doesn't exist there. Figure 3 depicts this situation. Point S on the object can be seen from the left camera but it cannot be seen from the right camera. So, if a window centered at point S in the left image is taken and a search is carried out in the right image a mismatch will be obtained.

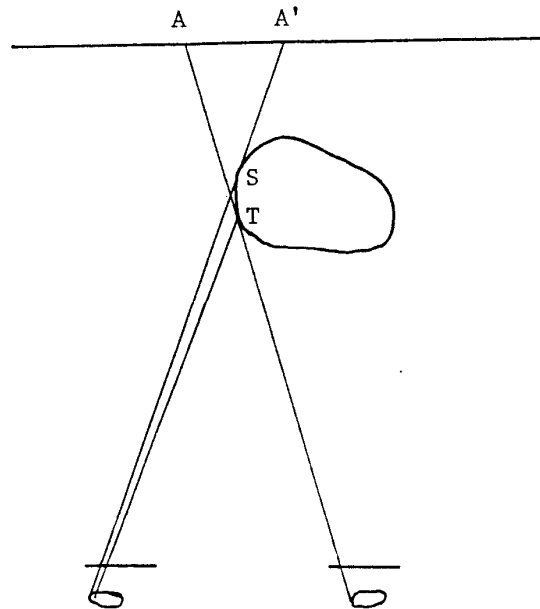


Figure 3. Image of an object in two stereo cameras.

To avoid mismatches caused by occluded points, we use the following fact. Referencing Figure 3, if point S on the boundary of a region in the left image does not exist in the right image, then on the same matching line, the point on the corresponding region in the right image (image of point T) can be seen in both of the images.

So, instead of selecting all points from one image for searching,

we select points from both images. On each match line, and for each corresponding region, we first take the window centered at the left image boundary and search it in the right image. Then take the window centered at the right image boundary and search it in the left image. Among the two matches, the one with the higher match rating is taken to be the true one. By this, if all mismatches due to occluded points are not recovered, their numbers will be reduced.

GEOMETRIC DIFFERENCES

With no doubt, windows centered at region boundaries contain more information than windows centered at points inside regions. This doesn't necessarily imply that matching on region boundaries is going to be absolutely error free. The major error that could cause a mismatch when searching along region boundaries is the existence of geometric difference between the images. Figure 3 depicts this fact. Note that background to the left of the object when viewed from the left camera corresponds to area A' while when viewed from the right camera corresponds to area A. Now if A and A' are considerably different, windows centered at region boundaries that should correspond to each other may not. Because the windows contain different parts from the background and this causes a low similarity measure for the windows, which consequently may cause a mismatch.

To avoid this, window shapes should be taken in such a way that they cover only parts from the object and not from both the object and the background. Figure 4 compares the traditional and the new window shapes.

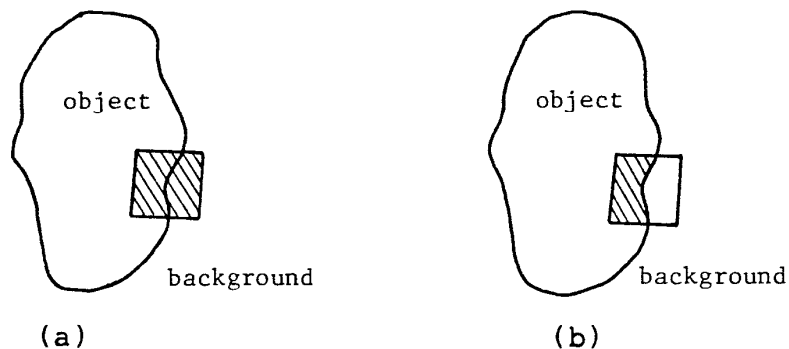


Figure 4. (a) Traditional and (b) new window shapes.

Implementation of the new window shapes is easy. Once the boundaries of regions are determined, the background can be replaced by zeros. When cross-correlation is computed, only non-zero pixels are used in the computation. Note that the

regions shouldn't contain zero-valued pixels (if they do, replace the background by a negative value and take non-negative values for determination of cross-correlation).

HOMOGENEOUS AREAS

If a selected window belongs to a homogeneous area, it won't contain enough information to distinguish it from other windows and so a mismatch might occur. Fortunately by taking windows centered at region boundaries the occurrence of low information windows become very rare. If selection of a homogeneous window is inevitable while following the boundary which might result in a mismatch, there are two ways to recover it.

- 1) If geometric models of objects in the scene are known and the objects in the images can be recognized, then fit the right geometric model to the boundary points which have high match ratings. By this, it is possible to detect and recover mismatches.
- 2) Since depths on a rigid object varies smoothly [11], if disparity of a point turns out to be different from its neighboring points by larger than a threshold value, then it is likely that a mismatch has occurred and so the disparity can be corrected using the disparities of neighboring points.

WEAKNESS OF THE SIMILARITY MEASURE

It has been experimentally shown that cross-correlation provides a higher accuracy than the sum of absolute differences as the window similarity measure [12]. Is cross-correlation the best similarity measure in window searching? There are other similarity measures like invariant moments and image transform coefficients that could as well be used. Since invariant moments or image transform coefficients are information preserving, if enough number of them are taken, the original windows can be reconstructed. It seems that invariant moments and image transform coefficients could provide a better similarity measure in the window search process.

The performance of cross-correlation versus invariant moments and image transform coefficients is not known. No attempt has been made to investigate it here either. However, we are considering using other similarity measures in the search process and compare their performances to cross-correlation. It might be that other similarity measures provide better accuracy.

RESULTS

To evaluate the performance of the described image disparity measurement technique, two experiments were conducted. In the first experiment, a three-dimensional scene was created by arranging a number of objects like a box, a reel of tape, a belt, a wad of paper, a piece of chalk, a roll of film, and a pipe cap, on a textured table. Figure 5 shows two images obtained at different viewpoints from the scene.

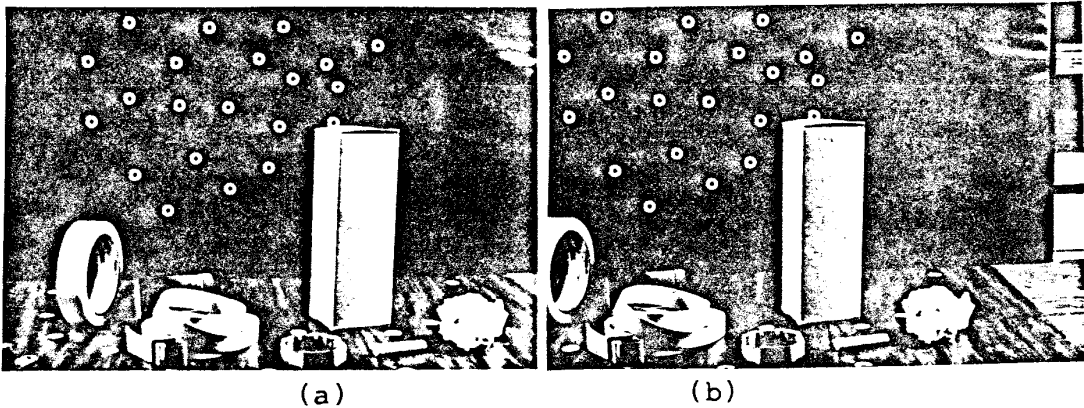


Figure 5. Stereo images of a three-dimensional scene.

Since a stereo camera system was not available, the images were obtained by the same camera with an appropriate displacement of the camera horizontally and parallel to the background. The background was marked with small spots so that the two images can be aligned by registering the background. Since by registering the background the disparity on the background becomes zero, position of points in the scene appear to be farther from the viewer than they should be.

The left hand side image was segmented and is shown in Figure 6 with region boundaries overlaid on the original image. Regions with perimeters smaller than 30 pixels were eliminated because they do not contain enough information for window matching. The background is replaced by zeros.

Window size 16x16 and search domain size 3x7 was selected. The reason for selecting 3 rows rather than 1 row for search is because a stereo camera has not been used to obtain the images. When putting the pictures down for digitization, a slight rotational difference between the images appears as if the camera's displacement hasn't been horizontal. And so two corresponding points in the two images might not fall on the same row. So, the neighboring rows also need to be searched for

elimination of this kind of error. It is assumed that neighboring points on the region boundaries do not have disparity difference of more than 3 pixels. So, 7 column places have been taken for search (3 pixels at either side of the previous corresponding point's column number).

To establish the first corresponding point on a boundary, a 5×33 search domain is selected. It has been assumed that maximum disparity in the images is 16 pixels. By this first step, a pair of corresponding points in the two images become known. Then the search domain is reduced to 3×7 for the rest of the points on the boundary.

Since sharp changes in depth causes the disparity of two neighboring points to differ by more than 3 pixels. If in the window search process the cross-correlation value for the matched windows fall below 0.6, the search is repeated. This time in a 3×17 area. The disparity of boundary points determined in this manner are shown in Figure 7. In this figure, the darker the point, the larger the disparity between the corresponding points in the two images.



Figure 6. Region boundaries overlaid on Figure 5.b.

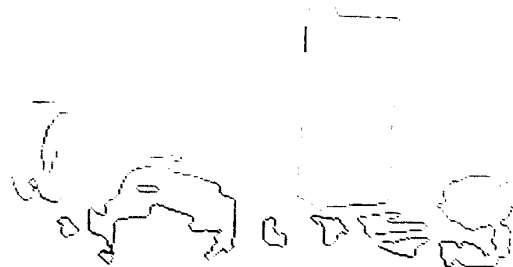


Figure 7. Disparity values on the region boundaries.

We use the property of adhesiveness of matter and the fact that depth changes smoothly on the surfaces of objects as follows. In the search process if the disparity of a point changes by more than 3 pixels with two of its immediate neighbors then the disparity of that point is replaced by the average disparity of the two neighboring pixels. By this, noisy disparities are corrected.

As it can be seen in Figure 6, some of the extracted regions contain more than one type of region. We have to recursively segment the regions until satisfactory regions are obtained (regions with variances below a threshold value). As an example,

the region corresponding to the box in Figure 6 contains two regions corresponding to two faces of the box. The two faces of the box have different intensities. The region corresponding to the box is extracted from Figure 6. This is shown in Figure 8. Now we assume Figure 8 is a new image and apply the segmentation technique to it.

If we overlay the boundary of the segmented box on the boundary of the originally segmented box, we obtain result of Figure 9. If we determine the disparity on the boundaries of the box and interpolate the disparity of points on the surfaces of the box using disparity of points on its boundaries, we obtain image of Figure 10.

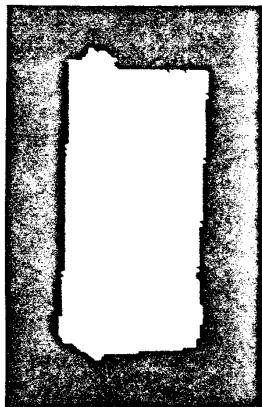


Figure 8.
Image of the Box.

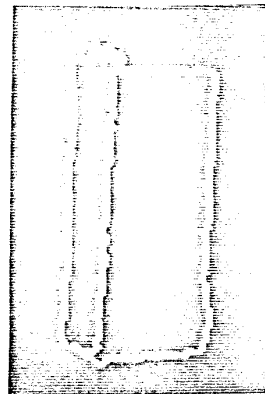
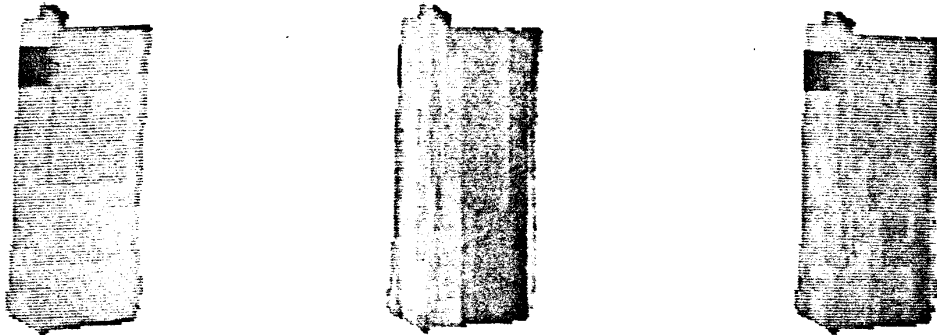


Figure 9.
Boundaries of the box.

As another example, two aerial photographs from a residential area as shown in Figure 11 were used. The right hand side image was segmented by level slicing the image manually so as to extract boundaries of buildings in the images. The result of segmentation is shown in Figure 12 with region boundaries overlaid on the original image and the background is removed.

Unlike the previous example where the arrangement of the scene and the lighting was controlled in the laboratory, in this experiment there has been no control over the scene or lighting conditions. There are different shadows in the two images and a large number of trees and bushes exist in the scene which make the disparity measurement difficult. Since views of trees, with shadows look very different from two different viewing angles, larger number of mismatches are obtained. As can be seen in Figure 13, mismatches have mainly occurred in areas where shadows and/or trees are present.

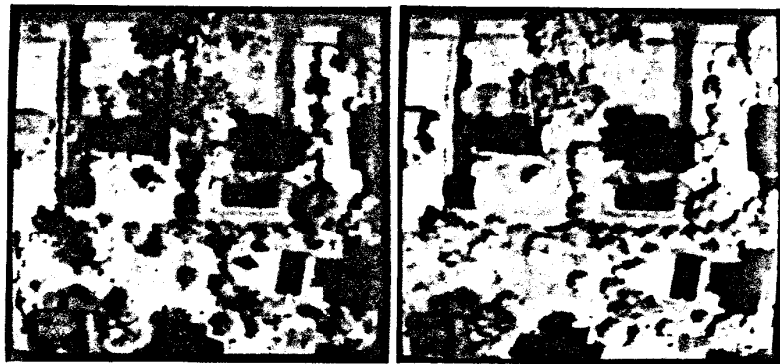
Since most extracted regions contained buildings with planar



(a) (b) (c)
 Figure 10. Disparity on faces of the box estimated (a) horizontally, (b) vertically, and (c) horizontally and vertically averaged.

surfaces, disparity inside regions were estimated by linear interpolation of disparity on region boundaries. This is shown in Figure 14. Note the two pools in the original image which could hardly be distinguished from buildings are classified correctly in the disparity map of Figure 14.

Finally, to show how much improvement is obtained by removing the background (taking window shapes such that they cover only parts from the object), the building at the middle-left of the scene in Figure 11 is selected. First, disparity on the boundary of the building was determined with background present, and then with background removed. Figures 15.a and 15.b show the computed disparities, correspondingly. In this experiment, when



(a) (b)
 Figure 11. Two stereo aerial photographs.

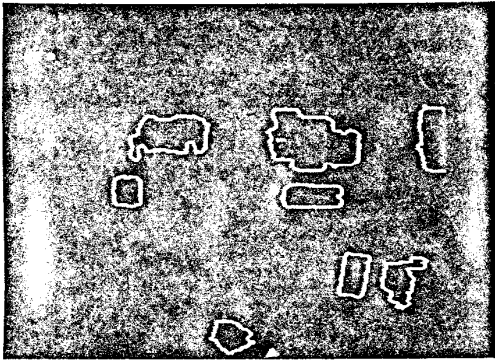


Figure 12. Segmentation of image of Figure 11.b.

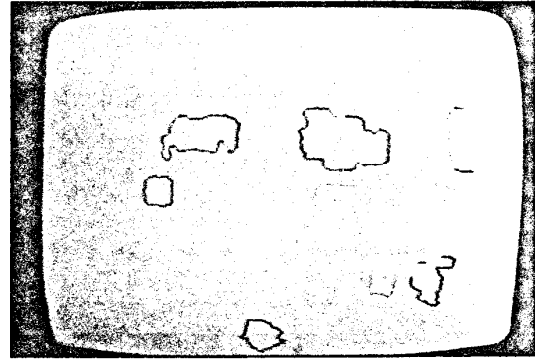


Figure 13. Disparity of points on the region boundaries.

background was removed, all disparities were computed correctly within a pixel accuracy (the disparities are multiplied by 15 for video display purposes), while when the background is present most of the disparities are wrong.

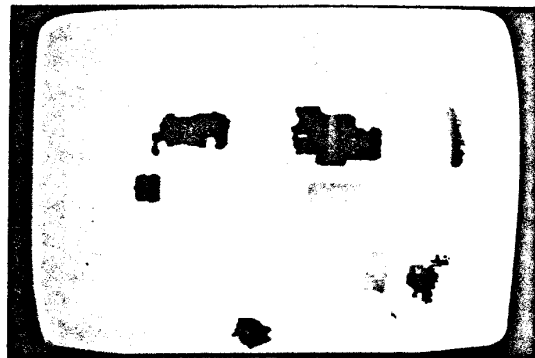


Figure 14. Estimated disparities inside regions.

CONCLUSIONS

Measurement of disparity for stereo depth perception was approached by first determining the disparity of points on region boundaries through window search and then determining disparity of other points in the image by interpolation or model fitting. Major problems causing the mismatches in the window search process were identified as 1) occluded points, 2) geometric differences, 3) homogeneous areas, and 4) weakness of the similarity measure. Appropriate counter measures were introduced to avoid occurrence of mismatches.

The proposed approach was tested with two sets of real data and improvement over the traditional disparity measurement via window searching was demonstrated.

Determination of a point disparity when cross-correlation coefficient for the best match is greater than or equal to 0.6, takes about 2 seconds and when cross-correlation coefficient for the best match is less than 0.6 takes about 18 seconds on a PDP 11/34 computer (where real addition time=65 micro seconds and real multiplication time=85 micro seconds). The number of points on region boundaries of Figures 7 and 13 are 1314 and 833, respectively. Total of 1 hour and 11 minutes for Figure 7 and 42 minutes for Figure 13 were required to determine the disparity on region boundaries.

REFERENCES

- [1] Marr, D. and T. Poggio, "A Computational Theory of Human Stereo Vision," Proc. R. Soc. Lon., B. 204, 1979, pp 301-328.
- [2] Grimson, W. E. L., "Aspects of a Computational Theory of Human Stereo Vision," Proc. Image Understanding Workshop, 1980, pp 128-149.
- [3] Baker, H. and T. O. Binford, "Depth from Edge and Intensity Based Stereo," 7th Int. Joint Conf. on Artificial Intelligence, 1981, pp 631-636.
- [4] Arnold, R. D., "Local Context in Matching Edges for Stereo Vision," Proc. Image Understanding Workshop, May 1978, pp 65-72.
- [5] Hannah, M. J., "Computer Matching of Areas in Stereo Images," Ph.D. Dissertation, July 1974, Computer Science Dept., Stanford University.

- [6] Moravec, H. P. "Rover Visual Obstacle Avoidance," 7th Int. Joint Conf. on Artificial Intelligence, 1981, pp 785-790.
- [7] Nevatia, R., "Depth Measurement by Motion Stereo," Computer Graphics and Image Processing 5, 1976, pp 203-214.
- [8] Davis, L. S., "A Survey of Edge Detection Techniques," Computer Graphics and Image Processing 4, 1975, pp 248-270.
- [9] Weszka, J. S. and A. Rosenfeld, "Histogram Modification for Threshold Selection," IEEE Trans. on Sys., Man, Cybern., Vol. SMC-9, Jan. 1979.
- [10] Zucker, S. W., "Region Growing: Childhood and Adolescence," Computer Graphics and Image Processing, 5, 1976, pp 382-399.
- [11] Marr, D. and T. Poggio, "Cooperative Computation of Stereo Disparity," Science, Vol. 194, Oct. 1976, pp 283-287.
- [12] Svedlow, M., C. D. McGillem, and P. E. Anuta, "Experimental Examination of Similarity Measures and Preprocessing Methods Used for Image Registration," Sym. Machine Processing of Remotely Sensed Data, 1976.