

A REFINED TECHNIQUE FOR STEREO DEPTH PERCEPTION

Ardeshir Goshtasby

Dept. of Computer Science, University of Kentucky

ABSTRACT

The problem with the present stereo depth perception techniques is that they cannot detect occluded points, they are sensitive to geometric and intensity difference between images, and homogeneous areas and shadows in the images make the matching process inaccurate. In this paper, a new technique is proposed that avoids selection of occluded points for matching, is resistant to geometric and intensity difference between images, and avoids matching of points from homogeneous areas and shadows for increased accuracy.

images. Global correspondence was used to resolve ambiguities among local correspondences.

Hannah-Moravec-Nevatia approached image point correspondence by window searching^{5,6,7}. In this approach, first a number of high variance windows are selected from one image. These windows were searched in the other image by a cross-correlation template-matching technique. Then centers of matched windows from the two images were taken as corresponding points.

I. INTRODUCTION

Images obtained at different viewpoints from a three-dimensional scene have geometric differences. This geometric difference is referred to as the disparity, and image disparity is the key factor in determination of depths of points in the scene.

The reliability of a depth perception system depends on the reliability of the matching process. In the following, first, main sources that cause the errors in a stereo matching process will be introduced and then a new matching technique that tries to avoid the errors will be presented.

II. MAIN SOURCES OF ERROR IN STEREO MATCHING

If the images are obtained by a well-balanced stereo camera system then image disparity is determined by the horizontal difference of corresponding points in the two images. If the images are obtained by one camera with displacement, then there is a need to align the two images before measuring image disparity. Image alignment is possible if a number of points in infinity in both images can be identified. Points in infinity in the two images are then registered, and positional difference between other corresponding points in the images determine the disparity values.

The main sources that cause the errors in matching of stereo images has been identified as occluded points, geometric and intensity difference between the images, homogeneous areas, shadows, and the employed matching technique⁸. In the following, each source is discussed in detail.

A. Occluded Points

Therefore, to be able to measure image disparity, we need to know the position of corresponding points in the images. The correspondence problem has been studied by different groups. Baker-Binford-Arnold determined corresponding points by matching high gradient edges in the images obtained by the Heuckel operator^{1,2}.

Occluded points are points in the scene that can be seen from only one of the cameras. If an occluded point is selected from one image and is searched in the other image, a mismatch will be obtained every time because the point does not exist in that image. Figure 1 depicts this situation. Point S on the object can be seen from the left camera but it cannot be seen from the right camera. Therefore, a search for point S in the right image will result in a mismatch.

B. Geometric and Intensity Differences

Marr-Poggio-Grimson matched non-horizontal zero-crossings in order to determine corresponding points³. Zero-crossings are defined as the second directional derivatives of intensities in an image and are used to locate intensity changes in the image. Zero-crossings were defined for four sizes of masks in reference 3. Zero-crossings from larger masks were used to establish global correspondence, while zero-crossings from smaller masks were used to find local correspondence in the

One of the major sources that cause the mismatches is the geometric and intensity difference between the images. In depth perception by window searching, if centers of two windows correspond to each other, then by going away from the window centers geometric and intensity difference between the windows increases. This

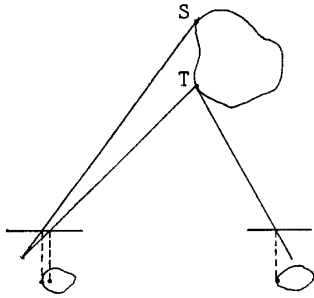


Figure 1. Images of an object in 3-D formed by a stereo camera system. Point S is occluded in the right image but point T can be seen in both images.

shows that pixels away from the window centers should be used with lesser importance than pixels near the window centers. Mori weighted the windows with Gaussian type weights to reduce the effect of distortions of points away from the window centers⁹. Taking note of the fact that smaller viewpoint differences introduce smaller geometric and intensity difference in stereo images, Nevatia used a sequence of images obtained at small viewpoint increments to reduce the amount of geometric differences⁷.

If matching is carried out using high-gradient edges or zero-crossings in the images, again geometric and intensity difference between the images could produce significantly different edges or zero-crossings causing the mismatches. Figure 2 shows an object in three-dimensions where the background of the object appears differently in two images. This image difference is the major source of error when matching points on the object boundaries whether or not window matching, edge matching, or matching of zero-crossings is used.

C. Homogeneous Areas

When window searching is used, if a selected window belongs to a homogeneous area, it may not contain enough information to distinguish it from neighboring windows and a mismatch may occur. Fortunately, when high gradient edges or zero-crossings are used for matching, homogeneous areas will not be used. However, there are cases where there is a need to determine depth of points on an object surface with homogeneous or slowly-varying intensity values. In such case, depth on the object surface can be determined by interpolation of depth of points on the object boundaries. For an excellent discussion of the interpolation problem see reference 13, Chapters 5 to 9.

Homogeneous areas in an image might be due to the fact that the camera has been out of focus. This happens often in stereo images because the depth of the scene is not known and therefore a focused image is not possible to obtain. However, using intensity distribution of an obtained image it is possible to interactively focus the cameras¹⁰.

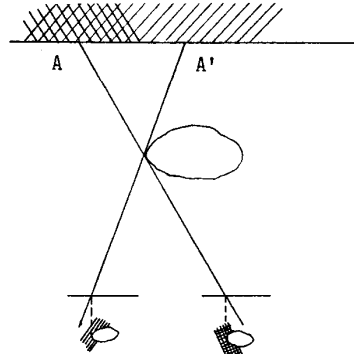


Figure 2. Images of an object in 3-D appear as if having different backgrounds.

D. Shadows

Shadows could produce considerable image difference especially in outdoor scenes. In an aerial stereo image pair, shadows of trees, bushes, and buildings appear differently in the two images. Matching of points in the shadows are not usually reliable because shadows mask the details in the scene. In the proposed technique of Section III, the images are segmented by a thresholding technique and since shadows usually fall in one side of the threshold, they are removed from the images.

E. The Matching Technique

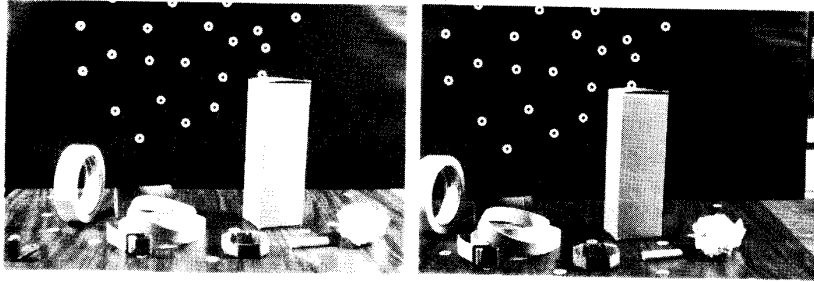
The employed matching technique has a significant role in determination of corresponding points. The edge-matching technique and the matching of zero-crossings are sensitive to geometric and intensity difference between the images while the window matching technique could be implemented to be less sensitive to geometric and intensity difference between images by weighting the windows with Gaussian type functions⁹. Other properties of the matching techniques that can affect matching accuracy are the size and number of masks used in the matching of zero-crossings, the choice of the edge operator in the edge-matching technique, and the choice of similarity measure and window size in the window-matching technique.

III. THE PROPOSED STEREO MATCHING TECHNIQUE

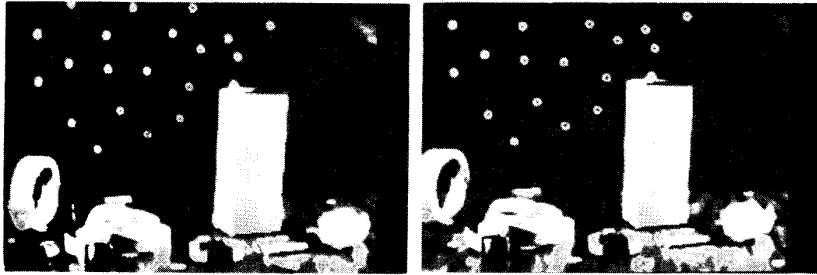
The proposed stereo matching technique consists of the following steps:

Step 1. Segment the two images such that similar regions are obtained. This segmentation technique has been described elsewhere^{8,11}.

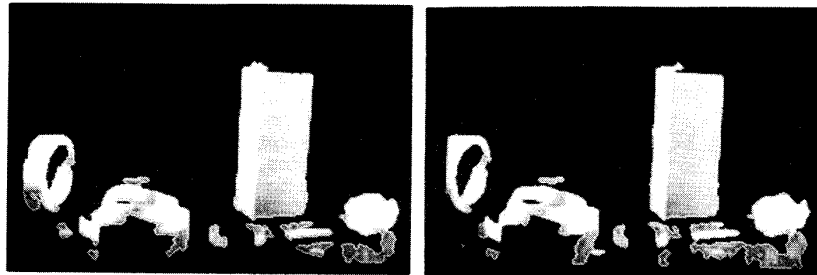
Figure 3.a shows stereo images obtained from a set of objects arranged on a textured table. The small circles on the background and on the table have been used to calibrate the cameras and to align the images. Figure 3.b shows images of Figure 3.a after segmentation and overlaying of the region boundaries on the original images.



(a)



(b)



(c)

Figure 3. (a) Original images of objects on a textured table. (b) Segmented images with boundaries overlaid. (c) Images with the background removed.

Step 2. Remove the background from the images.

When segmenting an image, two types of regions are obtained. The concept of object and background is arbitrary and either types of regions could be taken as the background. The remaining regions will correspond to the objects. However, to be able to remove the background that truly lays behind the objects, some a-priori knowledge about the scene is required. Figure 3.c shows images of Figure 3.b after removing the background and small objects (with perimeters less than 30 pixels) from the images. Removal of a region from an image could be thought of as replacing the region by zeros. Note that the remaining regions should not contain zero-valued pixels. Otherwise, they will be confused with the background in later steps of the

algorithm. If there are some zero-valued pixels in the non-removed regions, the background could be replaced by negative values.

Step 3. Determine corresponding regions in the two images. This step has also been described elsewhere^{8,12}.

In the following, it is assumed that the images have been obtained by a well-balanced stereo camera system so that corresponding points fall on the same row (called a match line) in the images. The original left image, the original right image, the left image with background removed, and the right image with background removed will be referred to as L, R, LB, and RB, respectively.

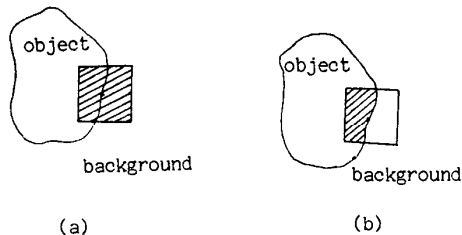
Step 4. For each region in LB where there is a correspondence in RB, perform steps 4.1 to 4.3.

Step 4.1. Determine the first pair of corresponding points on the given region boundaries.

To do this, the point on the region boundary of LB that has the smallest x-value is taken. A window centered at this point is selected and is searched in R on the same match line. The position giving the highest match rating is taken as the corresponding point. In this search, to determine similarity between two windows, non-zero values are used to avoid areas from the background. This act forces the shape of the window to fit the shape of the object. Figure 5 compares the traditional and the new window shapes. Since no part from the background is used in the matching, geometric and intensity difference from the background cannot affect the matching accuracy. If the maximum allowable disparity between the two images is d pixels, and the window of LB is centered at (x,y) , then there is a need to search along the same match line in R from $(x,y-d)$ to $(x,y+d)$. Since the selected point in LB might be occluded, the same process is repeated this time by taking the point with the smallest x-value from the region in RB and search it in L. Among the two matches, the one with the higher match rating is taken as the true one. This eliminates mismatches due to occluded points, see Figure 1. If image of point S on the region boundary of the left image cannot be seen in the right image, then on the same match line, the point on the region boundary of the right image (image of point T) can be seen in both images.



Figure 4. Computed disparities on region boundaries of Figure 3. The arrow shows the mismatches.



(a) (b)
Figure 5. (a) Traditional and (b) new window shapes.

Step 4.2. While following the region boundary in the clockwise (or counter-clockwise) direction, determine the next corresponding points. The search process for this step is similar to that of Step 4.1 except that the search domain is much smaller. Assuming the maximum difference in disparity of two neighboring points is d' , the search domain would be $2d'+1$.

Step 4.3. Repeat Step 4.2 until all corresponding points on the boundaries of the given region pair are determined.

Figure 4 shows the disparity values on region boundaries of images of Figure 3 using $d=20$, $d'=5$ and window size 16×16 . The darker the point, the larger the disparity. About 2% of the matches were incorrect which are shown by the arrow.

IV. CONCLUSION

Sources causing the errors in the presently available stereo matching techniques are: occluded points, geometric and intensity difference between images, homogeneous areas, shadows, and the applied matching technique. To avoid some of these errors, a refined matching technique was proposed with the following properties:

- 1) It tries to avoid occluded points by selecting points from both images for matching. In the traditional matching techniques, all points are selected from one image and are searched in the other image. In the proposed technique, points are selected from both images. For each match line and corresponding region boundaries, first the point from the left image is searched in the right image. Then the point from the right image is searched in the left image. Among the two searches, the one giving the higher match rating is taken as the true one. By this mechanism, selection of occluded points is avoided.
- 2) It tries to reduce the effect of geometric and intensity difference between the images by taking window shapes such that they cover only parts from the object and not the object and the background. To implement dynamically varying shaped windows that always fit the shape of region boundaries, the background is replaced by zeros, and when similarity between two windows is determined, non-zero values are used.
- 3) It avoids selection of points on homogeneous areas by selecting points on region boundaries where there usually are sharp changes in intensity.
- 4) Since one type of region is removed as the background, and since shadows in an image usually make the same type of region, it is possible to remove regions belonging to the shadows and therefore be able to avoid matching of points in the shadows.
- 5) The matching process is rather fast. Once a pair of corresponding points on corresponding region

boundaries are located, the correspondence between the rest of the points can be determined by following the region boundary and searching in limited areas. Determination of each corresponding point requires only $2d'+1$ matches. (In images of Figure 3, $d'=5$ was used.) The overall time required to segment images of Figure 3 (they are 240×240), to remove the background, to determine corresponding regions in the images, and to compute disparity on region boundaries takes 22 minutes on a PDP 11/34 computer where most of this time is spent on input-output.

ACKNOWLEDGEMENTS

The author would like to thank Drs. George Stockman and Carl Page of Michigan State University for their helpful suggestions and discussions on an earlier version of this work. The facilities at the Image Processing Laboratory of Michigan State University have been used to carry out experiments mentioned above. The help of Ms. Jane Spanyer in preparation of this paper is greatly appreciated.

REFERENCES

- [1] H. Baker and T. O. Binford, "Depth from Edge and Intensity Based Stereo," 7th Int. Joint Conf. on Artificial Intelligence, 1981, pp 631-636.
- [2] R. D. Arnold, "Local Context in Matching Edges from Stereo Vision," Proc. Image Understanding Workshop, May 1978, pp 65-72.
- [3] D. Marr and T. Poggio, "A Computational Theory of Human Stereo Vision," Proc. R. Soc. Lon., B. 204, 1979, pp 301-328.
- [4] W. E. L. Grimson, "Aspects of a Computational Theory of Human Stereo Vision," Proc. Image Understanding Workshop, 1980, pp 128-149.
- [5] M. J. Hannah, Computer Matching of Areas in Stereo Images, Ph.D. Dissertation, July 1974, Computer Science Dept., Stanford University.
- [6] H. P. Moravec, "Rover Visual Obstacle Avoidance," 7th Int. Joint Conf. on Artificial Intelligence, 1981, pp 785-790.
- [7] R. Nevatia, "Depth Measurement by Motion Stereo," Computer Graphics and Image Processing 5, 1976, pp 203-214.
- [8] A. Goshtasby, A Symbolically-Assisted Approach to Digital Image Registration with Application in Computer Vision, Ph.D. Dissertation, 1983, Dept. of Computer Science, Michigan State University.
- [9] K. Mori, M. Kidode, and H. Asada, "An Interactive Prediction and Correction Method for Stereo-comparison," Computer Graphics and Image Processing 2, 1973, pp 393-401.
- [10] R. A. Jarvis, "Focus Optimization Criteria for Computer Image Processing," MICROSCOPE 24, 1976, pp 163-180.
- [11] A. Goshtasby and C. V. Page, "A Multiple Image Segmentation Technique with Subpixel Accuracy," Proc. Computer Vision and Pattern Recognition, 1983, pp 157-158.
- [12] A. Goshtasby and C. V. Page, "Image Matching by a Probabilistic Relaxation Labeling Process," 7th Int, 1984. Conf. on Pattern Recognition.
- [13] E. L. Grimson, From Images to Surfaces: A Computational Study of the Human Early Visual System, MIT Press, 1981.