MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

A Stereo Vision through Creating a Virtual Image using Affine Transformation

Kimihiro UNO*
Department of
Electrical and Electronic Engineering
Yamaguchi University

Hidetoshi MIIKE†
Department of
KANSEI Design and Engineering
Yamaguchi University

Abstract

We propose a new approach for stereo vision. In human vision one percept can arise from two retinal images as a result of the process called "fusion". We try to create a virtual image from two images of stereo cameras. Optical axes of the cameras intersect on an attention point locating on the surface of an object. By Affine transformation of the respective images, the two images are transformed to a virtual image that is seen from the middle point between two cameras. By superimposing two transformed pictures, one fusion picture of virtual image is reduced. Excepting for around the attention point, two transformed pictures do not coincide with each other. Finding correspondence between two pictures, we can determine 3-dimentional depth information.

1 Introduction

Many approaches have been tested in stereo vision[1]~[5]. These are classified roughly into two categories. The first type, namely the intra-scanline search algorithm, is the traditional and popular method[1, 2]. Finding the correspondence of object points between a left and right image is the main problem in the approach. The depth information is then determined by triangulation. The second type is trying to model the human stereo system. On the basis of the psychophysics of human vision the human stereopsis theory is developed[3]~[5]. Especially, Griswold and Yeh proposed an interesting approach based upon the binocular fusion concept[6]. They pointed out several important findings. A crossed-looking type of camera model and dividing each image plane into right half and left half planes with respect to the center line are outstanding.

In this study we propose a new approach for stereo vision. According to Griswold and Yeh we introduce the knowledge established in neurophysiology and psychophysics, and we also refer to our own experiences on the binocular vision. As is well recognized, the two retinal images are unified in the visual cortex. The unification is achieved through the "superimposing" of their corresponding receptive fields. This is recognized as "binocular sensory fusion"[6]. In spite of many studies, however, no mechanistic explanation and a model and/or an algorithm to realize the unified image have not been proposed within our knowledge.

2 A Fusion Model for Stereo Vision

Our model for binocular vision is described as follows:

(1) A crossed-looking type of camera model is introduced as shown in Fig.1. Namely, optical axes of the cameras intersect on an attention point P_0 locating on the surface of the object.

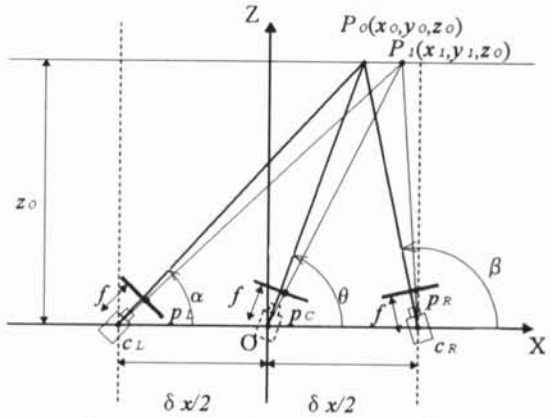


Figure 1: A geometric relationship for Affine transformation (Horizontal direction)

(2) An instantaneous fusion is limited only very small region around the attention point P_0. Continuous viewpoint shifting and rapid searching or saccade motion of eye play an important role to establish stereo vision.

* Faculty of Engineering, Yamaguchi University, Tokiwadai 2557, Ube 755, Japan.
E-mail: uno@sip.eee.yamaguchi-u.ac.jp
† Faculty of Engineering, Yamaguchi University, Tokiwadai 2557, Ube 755, Japan.
E-mail: miike@sip.eee.yamaguchi-u.ac.jp

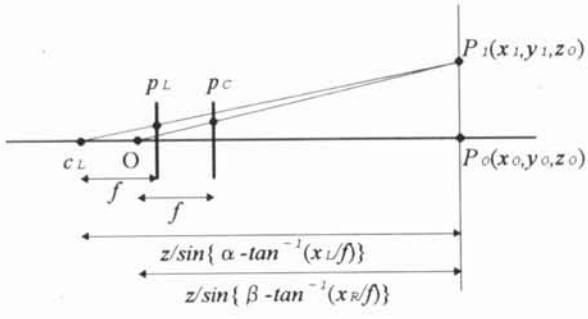


Figure 2: A geometric relationship for Affine transformation (Vertical direction)

(3) As the fusion model, the third virtual camera at the center position($X = 0$) between two real cameras(c_L and c_R) is introduced. Two images captured by left and right cameras are transformed into new pictures, which are assumed to observe the object from the center camera, by Affine transformation.

(4) In the transformation, the both attention points are projected to the center on each image plane. Textures on the respective images are assumed to lie on a plane parallel to a base-line connecting two cameras. Optical axes of the cameras are perpendicular to the plane(see Fig.2).

(5) The fusion can be achieved if the assumption is satisfied. Double vision on the superimposed virtual image shows that the assumption is not satisfied.

(6) Depth information at an arbitrary point P_2 can be determined by the disparity analysis. By solving the correspondence problem between two transformed pictures we obtain the depth map.

3 Theory

3.1 Affine transformation and the superimposed image

A schematic explanation to create the virtual image is shown in Fig.1 and Fig.2. The distance between an attention point $P_0(x_0,y_0,z_0)$ and the base-line(X-axis in Fig.1) is z_0. And the distance between two real cameras is δx. The focal length is f. The angles of respective camera are α(left camera), β(right camera) and θ(the center camera), respectively. Now we consider an arbitrary point P_1 neighboring P_0. When two points(P_1 and P_0) are on the same plane, which are parallel to the base-line (c_L-O-c_R), the following relationships can be obtained:

$$x_L = \frac{x_C(1+\tan^2\theta) f \tan^2\alpha}{(1+\tan^2\alpha)\tan\theta(f\tan\theta - x_C) + x_C(1+\tan^2\theta)\tan\alpha} \quad (1)$$

$$x_R = \frac{x_C(1+\tan^2\theta) f \tan^2\beta}{(1+\tan^2\beta)\tan\theta(f\tan\theta - x_C) + x_C(1+\tan^2\theta)\tan\beta} \quad (2)$$

$$y_L = \frac{\sin\left(\alpha - \tan^{-1}\frac{x_L}{f}\right)}{\sin\left(\theta - \tan^{-1}\frac{x_L}{f}\right)} y_C \quad (3)$$

$$y_R = \frac{\sin\left(\beta - \tan^{-1}\frac{x_R}{f}\right)}{\sin\left(\theta - \tan^{-1}\frac{x_R}{f}\right)} y_C \quad (4)$$

where, (x_L,y_L), (x_C,y_C) and (x_R,y_R) are the coordinates of point P_1 on the image planes of left real camera, the center virtual camera and right real camera, respectively. The equations (1) and (2) can be obtained by the horizontal geometric relationship(see Fig.1). The equations (3) and (4) can be obtained by the vertical geometric relationship(see Fig.2).

Through Affine transformation we obtain two transformed pictures. If the every texture points, in the original images, are lying on the plane parallel to the base-line(X-axis in Fig.1), the transformed pictures coincide with each other. A perfect fusion can be achieved in the case.

3.2 Detecting the depth map

However, because of 3D-shape of the object, the obtained textures are not lying on the same plane. Then, the superimposed image makes double vision. It implies that excepting for around the attention point two transformed images do not correspond with each other. Therefore, depth information at an arbitrary point P_2 can be determined by the disparity analysis as shown in Fig.3. Now we assume an interesting point P_2 is out of the plane. The point P_2 is projected onto p_L and p_R of the respective image planes on the real cameras(left and right). The distance between two cameras is δx. Then, the depth z_2 of P_2 can be determined by triangulation if the relationship between p_{C1}, p_{C2} and p_L, p_R are found. Where p_{C1} and p_{C2} are projections of point P_2 onto a image plane of the center camera. X-coordinates of p_{C1}, p_{C2}, p_L and p_R on the image plane are d_L, d_R, x_L and x_R, respectively. The results are summarized in the equations (5)-(7).

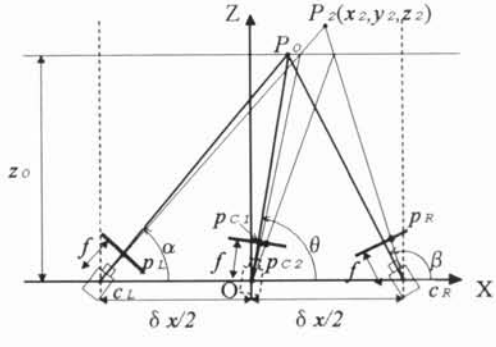


Figure 3: A geometric relationship for detection of the depth map

(a) The left image

(b) The right image

Figure 4: The real image pair

(a) The left transformed image

(b) The right transformed image

Figure 5: The transformed image pair

$$z_2 = \frac{\delta x}{\frac{f+x_L}{f\tan\alpha - x_L} - \frac{f+x_R}{f\tan\beta - x_R}} \quad (5)$$

$$x_L = \frac{f\{(\delta x\tan\alpha - 2z_0)(f\tan\theta - d_L) + 2z_0\tan\alpha(f+d_L\tan\theta)\}}{(2z_0\tan\alpha + \delta x)(f\tan\theta - d_L) + 2z_0(f+d_L\tan\theta)} \quad (6)$$

$$x_R = \frac{f\{(-\delta x\tan\beta - 2z_0)(f\tan\theta - d_R) + 2z_0\tan\beta(f+d_R\tan\theta)\}}{(2z_0\tan\beta - \delta x)(f\tan\theta - d_R) + 2z_0(f+d_R\tan\theta)} \quad (7)$$

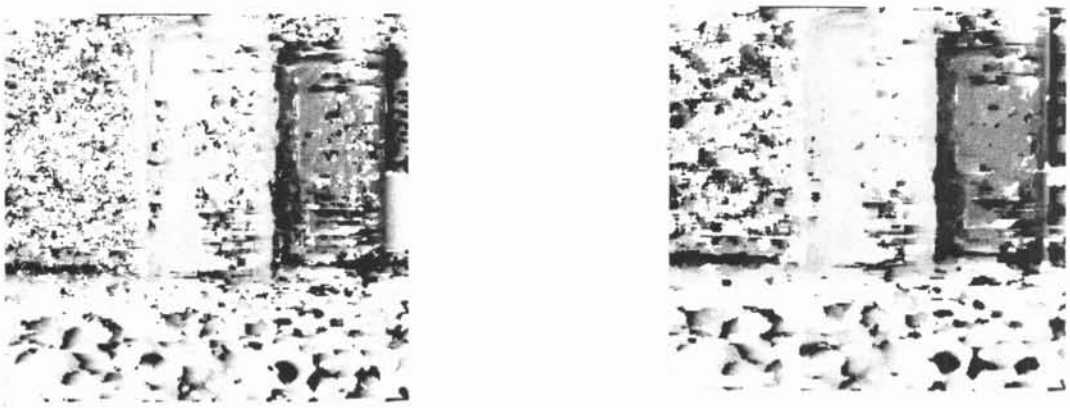
4 Experimental Results

We picked up an image pair in the following conditions.

The left camera angle: $\alpha \approx 88[^\circ]$
The right camera angle: $\beta \approx 94[^\circ]$
The distance between the camera: $\delta x \approx 5.5$[cm]
The distance of the object: $z_0 \approx 37.0$[cm]
The focal length: $f = 12.5$[mm]

In the first analysis, we created two new pictures, which are assumed to observe the object from the center camera, by the equations (1)-(4). And we took the summation operation of both transformed images. Now we obtain a virtual image of the superimposed picture. The real image pair is shown in Fig.4. And the transformed image pair is shown in Fig.5. The inside regions on transformed images were magnified, and the outside regions were reduced. Concretely, the right side on the left transformed image and the left side on the right image were magnified. And the left side on the left transformed image and the right side on the right transformed image were reduced. The superimposed picture is shown in Fig.6. There is the attention point on the center of the picture. The neighborhood of the attention point is clear, and the rest have double vision. The clear texture region in the picture guarantees a success of the fusion. The double vision region requires depth calculation.

Next step, we found correspondence between two transformed images with template matching(the size of the template is 3×3[pixel], and another size is 7×5[pixel]) and detected depth information. The results are shown in Fig.7.



(a) 3×3[pixel] template (b) 7×5[pixel] template

Figure 7: The depth map

Figure 6: The superimposed picture

5 Conclusion

Information of both eyes is fused inside the brain when the human being observes 3D-object. In this study, we propose a new concept of stereo vision. After Affine transformation of respective binocular images, the transformed pictures are superimposed in a virtual image. The virtual image represents a scene pictured from the center camera. Therefore, in general, the objects on the real image transformed into different size. This fact well represents the characteristics of human vision. When we switch our eye from monocular to binocular, we often recognize the change in the apparent size of the object. For 3D-object double vision is observed as shown in Fig.6, excepting for the attention point. The degree of discrepancy has information of depth. By introducing the template matching, we can reconstruct the depth map. Therefore, it seems true that the human being recognizes 3D-information in the brain by double vision, which is obtained from binocular real images.

Acknowledgments

The authors would like to thank the members of image and information processing laboratory and the other reviewers for their comments on this paper.

References

[1] H.Z.Dan and B.Dubuisson, *String Matching for Stereo Vision*, Pattern Recognition Letters 9, Elsevier Science Publishers B.V.,North-Holland, 1989, pp.117-126.

[2] Z.Zhang, *Estimating Motion and Structure from Correspondences of Line Segments Between Two Perspective Images*, IEEE Fifth International Conference on Computer Vision, 1995, pp.257-262.

[3] Y.Xiong and S.A.Shafer, *Hypergeometric Filters For Optical Flow and Affine Matching*, IEEE Fifth International Conference on Computer Vision, 1995, pp.771-776.

[4] A.L.Abbott and B.Zheng, *Active Fixation using Attentional Shifts, Affine Resampling, and Multiresolution Search*, IEEE Fifth International Conference on Computer Vision, 1995, pp.1002-1008.

[5] B.Boufama and R.Mohr, *Epipole and Fundamental Matrix Estimation using Virtual Parallax*, IEEE Fifth International Conference on Computer Vision, 1995, pp.1030-1036.

[6] N.C.Griswold and C.P.Yeh, *A New Stereo Vision Model Based upon the Binocular Fusion Concept*, Computer Vision, Graphics, and Image Processing 41, Academic Press, 1988, pp.153-171.