First Conference on Artificial Intelligence Applications
Denver, Colorado, Dec. 5-7, 1984

On Combining Stereopsis and Kineopsis for Space Perception

Amar MITICHE

INRS-Telecommunications, 3 place du Commerce, Verdun (P.Q.) H3E 1H6, Canada

0. Abstract

In this paper we show that optical flow can be integrated to the stereoscopic mechanism to determine three-dimensional motion. We also show that rigid motion can be characterized in such a way that objects with different motions in space can be easily segregated.

1. Introduction

Vision research has led to the emergence of *stereopsis* and *kineopsis* as the two principal views which explain some of the mechanisms of space perception.

Stereopsis, which is primarily supported by the work of Julesz ([1]) and some neurophysiological evidence ([2]), refers to the process of recovering depth from the retinal disparity field ([3]-[5]).

Kineopsis ([6]) refers to the process which operates on the optical flow pattern for space perception. Optical flow is the field of apparent velocities of images of object points in space, due to relative motion of viewing system and objects ([7]-[24]).

Stereopsis and kineopsis are generally believed to control different aspects of space perception and perhaps because of such atomistic views of space perception, stereopsis and kineopsis have been traditionally studied separately and the problem of combining their mechanisms or effects has been ignored. The question we ask here is : what can we gain from combining stereopsis and kineopsis that cannot be gained, or would be difficult to gain, from either stereopsis or kineopsis considered independently? We propose an answer by demonstrating the following :

(1) Optical flow can be integrated to the stereopsis mechanism to determine three-dimensional motion. To this effect, we will develop an analytical expression of three-dimensional motion in terms of depth, optical flow, and stereoscopy parameters.

(2) Once three-dimensional motion is determined, rigid motion and structure can be characterized and their perception achieved directly by a local test based on optical flow and depth. In some circumstances, this perception cannot be achieved using optical flow alone or depth alone.

Our approach, which could be appropiately called *stereokineopsis*, therefore provides an integrated computational process for the perception of space.

The remainder of this paper is organized as follows : Section 2 describes how three-dimensional motion can be obtained from depth, optical flow, and stereoscopy parameters. Section 3 treats the problem of perceiving rigidity from motion and Section 4 gives examples. Finally, Section 5 contains a summary.

2. Integrating Optical Flow to Stereopsis for Motion

We will adopt the configuration of cameras illustrated in Figure 1 (the origin is at the projection center and the focal lengths are unity).

Let P be a point in space, with coordinates (X_1, Y_1, Z_1) in S_1 and (X_2, Y_2, Z_2) in S_2. Let the image of P be p_1 on I_1 with coordinates (x_1, y_1) and p_2 on I_2 with coordinates (x_2, y_2). Finally, let the transform taking S_1 onto S_2 be decomposed into a rotation R about an axis through the origin O_1, and a translation T. The rotation is described by the matrix :

$$R = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

and the translation by the vector $T = (X_T, Y_T, Z_T)$. With these notations we can write :

$$(X_2, Y_2, Z_2) = (X_1, Y_1, Z_1).R \ + \ T$$

or, in expanded form

$$X_2 = a_{11}X_1 + a_{21}Y_1 + a_{31}Z_1 + X_T$$
$$Y_2 = a_{12}X_1 + a_{22}Y_1 + a_{32}Z_1 + Y_T$$
$$Z_2 = a_{13}X_1 + a_{23}Y_1 + a_{33}Z_1 + Z_T$$

Since

$$x_2 = \frac{X_2}{Z_2} \quad ; \quad y_2 = \frac{Y_2}{Z_2}$$

we have

$$x_2 = \frac{a_{11}X_1 + a_{21}Y_1 + a_{31}Z_1 + X_T}{a_{13}X_1 + a_{23}Y_1 + a_{33}Z_1 + Z_T}$$

Dividing the numerator and denominator of the right-hand side by Z_1, we obtain

$$x_2 = \frac{a_{11}x_1 + a_{21}y_1 + a_{31} + \frac{X_T}{Z_1}}{a_{13}x_1 + a_{23}y_1 + a_{33} + \frac{Z_T}{Z_1}}$$

Similarly for y_2

$$y_2 = \frac{a_{12}x_1 + a_{22}y_1 + a_{32} + \frac{Y_T}{Z_1}}{a_{13}x_1 + a_{23}y_1 + a_{33} + \frac{Z_T}{Z_1}}$$

We can therefore write x_2 and y_2 as functions of x_1, y_1 and Z_1. Thus, knowing x_1 and y_1 we can determine x_2 and y_2 from depth and camera calibration parameters. We adopt the following notation for the above relations, showing now dependence on time :

$$x_2(t) = \alpha(x_1(t), y_1(t), Z_1(t))$$

$$y_2(t) = \beta(x_1(t), y_1(t), Z_1(t))$$

But because optical flow at p_2 is

$$u_2 = \frac{dx_2}{dt} \quad ; \quad v_2 = \frac{dy_2}{dt}$$

we can write, where δ designates the partial, and d the total, differential :

$$\begin{aligned} u_2 &= \frac{\delta\alpha}{\delta x_1}.\frac{dx_1}{dt} + \frac{\delta\alpha}{\delta y_1}.\frac{dy_1}{dt} + \frac{\delta\alpha}{\delta Z_1}.\frac{dZ_1}{dt} \\ &= \frac{\delta\alpha}{\delta x_1}u_1 + \frac{\delta\alpha}{\delta y_1}v_1 + \frac{\delta\alpha}{\delta Z_1}.\frac{dZ_1}{dt} \end{aligned} \qquad (1)$$

where (u_1, v_1) is the optical flow at p_1. Similarly for v_2 we have

$$v_2 = \frac{\delta\beta}{\delta x_1}u_1 + \frac{\delta\beta}{\delta y_1}v_1 + \frac{\delta\beta}{\delta Z_1}.\frac{dZ_1}{dt} \qquad (2)$$

We see that if points p_1 in I_1 and p_2 in I_2 are in correspondence, then their velocities are related by the above equations. If this correspondence is established and stereoscopy parameters are known then functions $\alpha(t)$ and $\beta(t)$ and their derivatives appearing in Equations (1) and (2) are completely specified. We can then determine $Z_1' = \frac{dZ_1}{dt}$ from either Equation (1), which involves the horizontal components of optical flow, u_1 and u_2 and the vertical component v_1, at p_1 and p_2 or Equation (2), which involves the vertical components of optical flow, v_1 and v_2 and the horizontal component u_1, at p_1 and p_2.

Once Z_1' is computed, we can obtain the other two components of three-dimensional velocity as follows. Since

$$X_1 = x_1 Z_1$$
$$Y_1 = y_1 Z_1$$

Then

$$X_1' = x_1 Z_1' + u_1 Z_1$$
$$Y_1' = y_1 Z_1' + v_1 Z_1$$

Example : If S_1 and S_2 differ by a translation along the X −axis, then it is easy to verify that :

$$Z_1' = \frac{u_1 - u_2}{X_T} Z_1^2$$

The preceding analysis is summarized as follows:

Proposition : If correspondence is established between two image points p_1 and p_2 of a binocular view, then the three-dimensional motion of space point P *that they represent is known, given the stereoscopy parameters and the optical flow at p_1 and p_2.*

3. Perception of Rigid Objects in Motion

In the previous section we have shown how three-dimensional motion of points in space can be recovered from depth and optical flow. In this section we would like to show how perception of rigid objects in motion can achieved. We will use a characterization of rigid motion in the form of a theorem which relates depth and optical flow to the motion of rigid objects in space. This theorem (a reformulation derived in [24] of a classic theorem in mechanics [25]) states that a motion is that of a rigid object, E, if and only if :

$$(\forall t)(\forall P_1 \in E)(\forall P_2 \in E)$$
$$aZ_1^2 + bZ_2^2 + cZ_1Z_2 + dZ_1Z_1' + eZ_1Z_2' + fZ_2Z_2' + eZ_2Z_1' = 0 \quad (3)$$

Where t is time, Z designates depth and Z' the Z-component of three-dimensional velocity (the derivative of depth with respect to time); P_1, P_2 have image coordinates (x_1, y_1), (x_2, y_2) and optical flow values (u_1, v_1), (u_2, v_2) respectively and

$$a = -u_1x_1 - v_1y_1$$
$$b = -u_2x_2 - v_2y_2$$
$$c = u_1x_2 + v_1y_2 + u_2x_1 + v_2y_1$$
$$d = -x_1^2 - y_1^2 - 1$$
$$e = x_1x_2 + y_1y_2 + 1$$
$$f = -x_2^2 - y_2^2 - 1$$

The rigid motion theorem in the form above relates optical flow and depth explicitly to three-dimensional motion. Note that coefficients a, b, c, d, e and f are in terms of image positions and optical velocities flow values only. One can readily see that this theorem is, in fact, a test for rigidity that will discriminate between objects in different motions. Indeed, the test is expected to fail at motion boundaries and the extent of rigid regions will therefore be delinated. More precisely, given the value of image variables (image positions and optical flow) and the value of object variables (depth and velocity in depth) corresponding to a pair of points in the image, one can quickly check if these points are indeed the projections of points on the same rigid object in space using the above theorem.

Although the theorem refers to all pairs of points of a body of points, in practice we will assume that local rigidity implies global rigidity such that it is enough to apply the test to pairs of points in small neighborhoods.

Note that, because it uses both depth and optical flow, the test above should be inherently more powerful than those which use depth alone or optical flow alone. In the next section we will give examples that will illustrate this fact and show the expected capabilities of the test.

4. Example

In the following example, the camera positions differ by a translation along the X-axis. The scene of interest is illustrated in Figure 2. In this, an object is made of a rigid collection of planar patches A, B, and C. We will refer to this object simply as the object. Object D is a planar surface which plays the role of a background. We will refer to it as the background. A snapshot of this scene is taken from viewpoint O, the viewing direction being the Z-axis. Refer to Figure 3. At the instant of the snapshot, the geometric configuration is such that D is at $Z = 4m.$, A at $Z = 2m.$ and B is at $Z = 3m.$ (C is not visible and does not play a role here. The purpose of including it in this example is to have a rigid object physically meaningful). The motion of the object is a rotation about the Z-axis (the angular velocity vector is $(0, 0, 2)$) and that of the background is a screw composed of a translation (the translational velocity is $(0, 0, 1)$) and rotation (the angular velocity is the same as that of the object $(0, 0, 2)$).

We are looking at a portion of the image around the origin. Before attempting to segment such a scene, we added noise to depth as well as to optical flow. Noise is uniformly distributed between $-\alpha a$ and $+\alpha a$ where a is the value to which noise is added and α is a given fraction of this value. In this particular example $\alpha = .5$ for depth and $\alpha = 1.5$ for optical flow. The average absolute value of error is therefore 25% for depth and 75% for optical flow(1). The result of applying the rigidity test(2), as described in the previous section, is shown in Figure 4. Pluses indicates motion boundaries, i.e., points where this test has failed. The

(1) This amount of noise is at the limit of what the method can do on this particular example without making a mistake in the segmentation. Sensitivity to noise will in general depend on the geometry of the scene and the motions involved.

(2) In these examples, the test is performed by comparing the absolute value of the left-hand side of (3) to a small threshold. We have not addressed the problem of automatic threshold selection. The test has been applied to each point and its right, lower, and lower diagonal neighbors.

scene has been correctly segmented and we can make the following important observations :

(1) If we had to segment the scene from depth alone then, because of the relashionship between the depths of regions A, E, and D (B is at the same distance from A and C), there is no way one can find a threshold on depth that would consistently declare A and B part of the same rigid object without also including the background D.

(2) As for this example, it can be shown that the contribution of translation to optical flow values is very small because of perspective around the origin. Since both the background and the object have the same rotational velocity (their motion differ in the translational velocity), it is practically impossible to discriminate between them on the basis of optical flow alone.

5. Summary

In this study we have addressed the problem of combining two important sources of three-dimensional information : stereopsis and kineopsis. We have described a method in which optical flow was integrated to the stereopsis mechanism to determine three-dimensional motion. Then motion of rigid objects was characterized and this characterization was the basis for a simple test that allowed the segregation of objects with different motions in space. This test, which used both optical flow and depth, was shown to be more powerful than those which use optical flow alone or depth alone.

REFERENCES

1. Julesz, B., '*Foundations of Cyclopean Perception*', University of Chicago Press, Chicago, 1971.

2. Barlow, H.B., Blakemore, C., and J.D. Pettigrew, 'The Neural Mechanism of Binocular Perception', *J. Physiol.*, 193, 327-342, 1967.

3. Aggarwal, J.K., Davis, L.S., and W.N. Martin, 'Correspondence Processes in Dynamic Scene Analysis', *Proc. IEEE*, 69, no 5, 562-572, May 1982.

4. Marr, D., and T. Poggio, 'A Computational Theory of Human Stereo Vision', *Proc. R. Soc. Lond.*, B.204, 301-328, 1979.

5. Duda, R.O., and P.E. Hart, '*Pattern Classification and Scene Analysis*, Wiley-Interscience, 1973.

6. Nakayama, K., and J.M. Loomis, 'Optical Velocity Patterns, Velocity-sensitive Neurons, and Space Perception: a Hypothesis', *Perception*, 3, 63-80, 1974.

7. Fennema C.L., and W.B. Thompson, 'Velocity Determination in Scenes Containing Several Moving Objects' *Computer Graphics and Image Processing*, 9, 301-315, 1979.

8. Horn, B.K.P., and B.G. Schunk, 'Determining Optical Flow', *Artificial Intelligence*, 17, 185-203, 1981.

9. Paquin, R., and E. Dubois, 'A Spatio-temporal Gradient Method for Estimating the Displacement Field in Time-Varying Imagery', *Computer Vision, Graphics, and Image Processing*, 205-221, 1983.

10. Barnard, S.T., and W.B. Thompson, 'Disparity Analysis in Images', IEEE *Trans. Pattern Analysis and Machine Intelligence*, 2, 333-340, 1980.

11. Nagel, H.H., 'On Change Detection and Displacement Estimation in Image Sequences', *Pattern Recognition letters*, 1, 55-59, 1982.

12. Nagel, H.H., 'Displacement Vectors Derived from Second-Order Intensity Variation in Image Sequences', *Computer Vision, Graphics, and Image Processing*, 21, 85-117, 1983.

13. K. Wohn, L.S. Davis, and P. Thrift, 'Motion Estimation Based on Multiple Local Constraints and Nonlinear Smoothing', *Pattern Recognition*, 16, No 6, pp 563-570, 1983.

14. Bruss, A.R., and B.K.P. Horn, 'Passive Navigation', *Computer Vision, Graphics, and Image processing*, 21, 3-20, 1983.

15. Lawton, D.T., and J.H. Rieger, 'The Use of Difference Fields in Processing Sensor Motion', Proc. Image Understanding Workshop, Washington D.C., 77-83, 1983.

16. Longuet-Higgins, H.C., and K. Prazdny, 'The Interpretation of a Moving Retinal Image', Proc. *R. Soc. Lond.*, B208, 385-397, 1980.

17. Prazdny, K., 'On the Information in Optical Flows', *Computer Vision, Graphics, and Image*

Processing, 22, 239-259, 1983.

18. Waxman, A.M., 'Kinematics of Image Flow', Proc. Image Understanding Workshop, Washington D.C., 175-181, 1983.

19. Gibson, E.J., Gibson J.J., Smith O.W., and H. Flock, 'Motion Parallax as a Determinant of Perceived Depth', *J. of Experimental Psychology*, 58, 40-51, 1959.

20. Thompson, W.B., and S.T. Barnard, 'Lower-Level Estimation and Interpretation of Visual Motion', *Computer*, 14, 20-28, 1981.

21. Thompson, W.B., 'Combining Motion and Contrast for Segmentation', IEEE *Trans. on Pattern Analysis and Machine Intelligence*, 2, 543-549, 1980

22. Potter, J.L., 'Scene Segmentation Using Motion Information', *Computer Graphics and Image Processing*, 6, 558-581, 1972.

23. Ballard, D.H., and O.A. Kimball, 'Rigid Body Motion from Optical Flow and Depth', *Computer Vision, Graphics, and Image Processing*, 22, 95-115, 1983.

24. Mitiche, A., 'Computation of Optical Flow and Rigid Motion', 2nd *IEEE* Workshop on Computer Vision : Representation and Control, April 30-May 2, Annapolis, MD, 1984.

25. Lelong-Ferrand, J., and J.M. Arnaudiès, 'Cours de Mathématiques - Tome 3, Géométrie et Cinématique', Dunod, Paris, 1974.

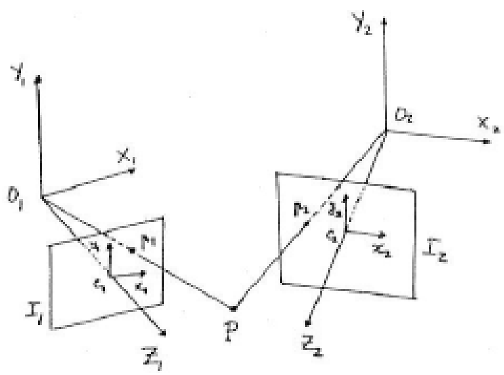

Figure 1

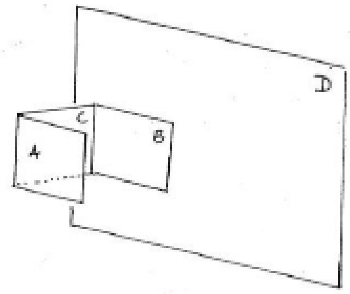

Figure 2

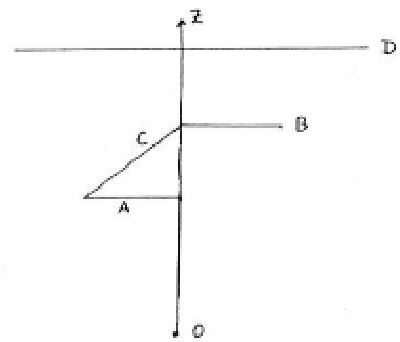

Figure 3

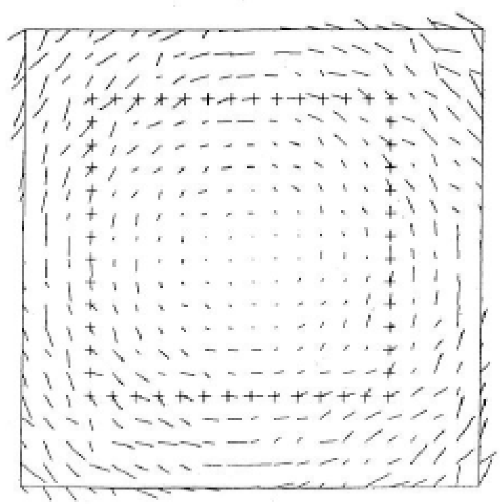

Figure 4