STEREO CORRESPONDENCE FROM OPTIC FLOW

Valérie Cornilleau-Pérès and Jacques Droulez
Laboratoire de Physiologie Neurosensorielle, CNRS
15 rue de l'Ecole de Médecine, 75270 Paris cedex 06, France

The cooperation between motion and stereopsis in the perception of 3-D structure has received several contributions [4,6,7,8]. These studies are motivated by the similarity between the algorithms of motion processing and stereopsis (the 2-D correspondence between image points feeds a 3-D interpretation stage), and by the need for reducing their sensitivities to noise. Most of them use both the optical flow and the intensity-based correspondence between each pair of stereo images as the inputs for the search of 3-D parameters. In [10] it is shown that motion and stereopsis can cooperate earlier, namely during the processing of the correspondence between stereo images. However this approach concerns only rigid surfaces of small curvature (locally approximated by their tangent plane), and its robustness under noise was not tested. In this paper, we show that the retinal velocity can be used directly for stereo matching, for any type of 3-D environment. The case of rigidity is used for predicting the performance of the method. We then present some computer simulations, and finally discuss the relevance of **stereo correspondence from optic flow (SCOF)** for biological vision.

1. The theoretical approach to SCOF.

Let us consider two optical systems of centers O_1 and O_2, with coplanar optical axes. The projection surfaces S_1 and S_2 are hemispheric, of centers O_1 and O_2. O is the middle of the baseline $[O_1O_2]$. In the orthogonal system (OIJK), I is a vector of the line (O_1O_2), and K is in the plane of the 2 optical axes. The coordinate system (OIjk) is obtained from (OIJK) by a rotation of angle $\varnothing$ of J and K around (OI) (Fig.1a), j is chosen unitary.

The points M_2 of S_2 that are possible correspondents of a point M_1 of S_1 lie on the epipolar line $E(M_1)$, which is the intersection of S_2 with the plane $(O_1O_2M_1)$. Given M_1 on S_1, there is a unique angle $\varnothing$ such that the plane (OIk) contains M_1, O_1 and O_2, and the search for the correspondent M_2 to M_1 is restrained to this plane which is the plane of Fig.1b.

Θ_1 is the angle between (O_1O_2) and (M_1O_1). In the orthogonal coordinate system $(O_1i_1jk_1)$, k_1 is a vector of the line (M_1O_1). Θ_2, i_2 and k_2 are defined accordingly for M_2 (Fig.2).

Let M be an object point moving with a 3-D velocity U, P_1 and P_2 the reciprocal of the distances O_1M and O_2M respectively, v_1 the velocity component along j of the image M_1 of M on the left image. The same notations hold for M_2, image of M on S_2. If $< , >$ refers to the inner product, we have: $v_1 = P_1.<U, j>$ and $v_2 = P_2.<U, j>$.

Considering that $P_1.\sin\Theta_2 = P_2.\sin\Theta_1$, it implies that:

$$v_1/\sin\Theta_1 = v_2/\sin\Theta_2 \quad (1)$$

Given M_1 on S_1, any point M_2 of $E(M_1)$ verifying eq. (1) is a possible correspondent of M_1. More generally, for any pair of viewing systems of optic centers O_1 and O_2, the *static epipolar constraint* expresses that correspondent points are coplanar with O_1 and O_2, while the *dynamic epipolar constraint*, in the form of eq. (1) here, *states that this coplanarity must be preserved during motion.* Therefore stereo-correspondence from optical flow (SCOF) consists in the pairing of points that satisfy both the static and dynamic epipolar constraints. SCOF is then equivalent to the classical process of stereo matching, where the luminous intensity is replaced by the function f defined along an epipolar line as: $f(\Theta_i) = v_i/\sin\Theta_i$ (i=1 or 2). If the noise which perturbates the measure of velocity is proportional to velocity-magnitude, the performance of SCOF must increase with the ratio $f'(\Theta)/f(\Theta)$ (we drop index i, and f' is the Θ-derivative of f along an epipolar line).

The case of rigidity. For a rigid surface, the motion can be decomposed in a rotation around O and a translation, of coordinates (w_X, w_Y, w_Z) and (t_X, t_Y, t_Z) in (IJK), and we have

$$f(\Theta) = [t_Y.\cos\emptyset - t_Z.\sin\emptyset + a.(w_Y.\sin\emptyset + w_Z.\cos\emptyset)]/z - (w_Y.\sin\emptyset + w_Z.\cos\emptyset)/tg\Theta - w_X \quad (2)$$

where z is the k-coordinate of M and $\emptyset$ the vertical excentricity of M. Eq. (2) indicates that SCOF is not always possible (for instance if w_Y and w_Z are null and the surface is such that z does not vary with Θ), and that the performance of SCOF should be optimal when the depth variations are large in the direction of (O_1O_2). Finally t_X and w_X are useless for the SCOF, but w_X should play a role for noisy data, when the noise is proportional to |f|.

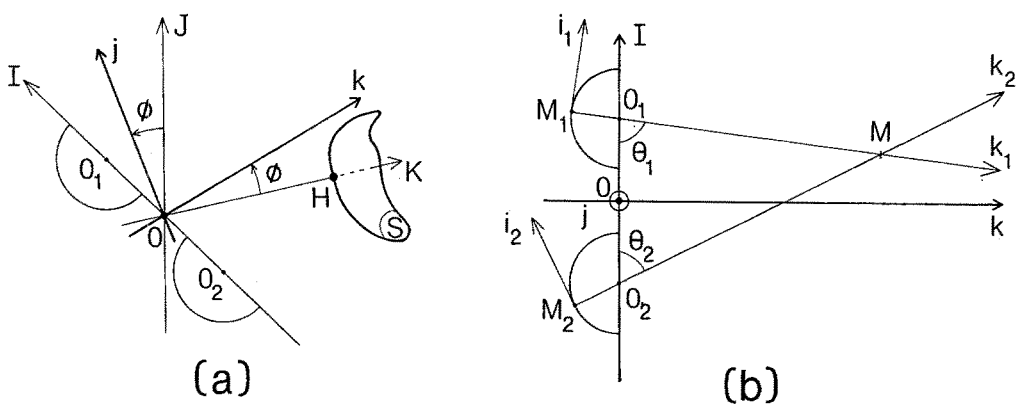

Figure 1. The coordinate systems.

(a) O_1 and O_2 are the optical centers of the left and right viewing systems. O is the middle of $[O_1O_2]$, (OK) is orthogonal to the line (O_1O_2) represented by the axis (OI). The axis J forms an orthogonal system with I and K. The system (OIjk) is obtained from (OIJK) by a rotation of angle $\emptyset$ around (OI). The 2 arcs of circles represent the hemispherical retinae. S is a surface intersecting (OK) in point H.

(b) M is an object point of images M_1 and M_2 on the left and right retinae respectively. The local referential ($M_1i_1jk_1$) and ($M_2i_2jk_2$) and the angles Θ_1 and Θ_2 are defined in text.

2. Computer simulations.

The first computations concern foveal vision, when the influence of each motion component is uniform over the image ($|\Omega|$ small), while the second group handles large planar images where the influence of motion components varies over the image.

Initially the surfaces intersect the axis (OK) in a point H, 72 cm from O (Fig.1a). The *time unit tu* is arbitrary. The 3-D velocities are chosen so as to limit the image velocity to 1 image diameter per tu. Different types of motions are used; among them, motion *R* is a rotation around the axis (HJ) (i.e., a translation along (OI) associated to a rotation around (OJ)), and motion *O* is a rotation around (OK).

The velocity coordinate v along j is perturbated by 2 Gaussian noises, one with standard deviation proportional to |v| (0.03|v|), the other of constant standard deviation (1 pix/tu). The noisy velocity $\underline{v}$ is filtered with a 11X11 pix Gaussian of radius 3 pix. A linear interpolation is then applied between epipolar lines. Given a pixel p_l of the left image, the algorithm searches for consecutive pixels p_r and p_r' in the right epipolar such that $f(p_l)$ ranges between $f(p_r)$ and $f(p_r')$ (with straightforward notations). p_l is termed 'high-confidence point' if it satisfies two conditions: (i) $|f(p_l)|$ is higher than 4 times the background noise; (ii) the variation of f between the two nearest neighbours of p_l exceeds 1%. When (ii) is not verified, or when no correspondent to p_l is found, the depth is set to that found in the previous epipolar pixel. The absolute depth is calculated by triangulation and the error thus performed, relative to the theoretical depth, is averaged over the high-confidence points. We also plot the depth profile in the plane (OIK) (horizontal profile).

Small field simulations. The field of view is 8° diameter (40 pixels). As $|\Omega|$ remains small, eq. (2) indicates that t_Z and w_Y play a minor role relative to t_Y and w_Z, and motions *R* and *O* represent the two basic motions to be tested in this condition.

For motion R (Fig. 2a,b) the performance of the algorithm improves as the surface is more tilted in the direction of (O_1O_2) (75% of high-confidence points when the tilt-angle reaches 10°). For different surfaces the mean depth error is less than 1% whenever more than 30% points are matched confidently. *For motion O* (Fig. 2c,d) more than 73% of the points could be matched confidently, with a mean depth error smaller than 0.7%. These two motions yield a minimal retinal slip (null velocity in the center of the image). This condition was found to be critical here; for instance a 12% reduction of w_X in motion *R* impaired greatly the performance.

Large field simulations. The 120X120 pix images cover 60° viewing angle. For *motions R and O* (Fig. 3) the results are the same as for small field simulations, except that the depth-error is 8 to 12 times higher (this being partly due to the loss of resolution). For a wide variety of surfaces and motions, when the high-confidence area is larger than 70%, the depth

error ranges generally between 1.5 and 15%. The algorithm was also much less sensitive to an increase of the retinal slip (without change of the maximal image velocity) than in the case of small viewing angle.

Overall, our results indicate that the robustness of SCOF to the type of noise that we used is optimal when w_X minimizes the overall image velocity. For any surface, motion *O* (rotation around the sagittal axis (OK)) provides reliable velocity information for the SCOF which then presents robustness under noise. In addition, motion *R*, as well as pure translations along (OK) for large viewing angles, allow a good reconstruction of the parts of the environment with steep variations in depth.

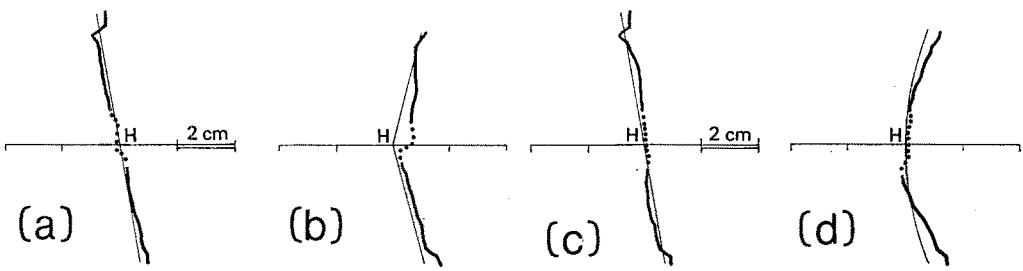

Figure 2. Results of small field simulations. The plane of the figure is (OIK) and the graduated axis is (OK). The thin line indicates the exact section of the surface, while the reconstructed section is in thick line. The dashed line shows the low-confidence points (see text). Point O is located 72 cm from H in the left direction of the figure. The scale is identical for all pannels in the horizontal and vertical directions. (a) Motion *R*, plane of tilt angle 10° in the direction (O_1O_2). (b) Motion *R*, 160° dihedron. (c) Motion *O*, same plane as in (a). (d) Motion *O*, sphere of radius 10 cm.

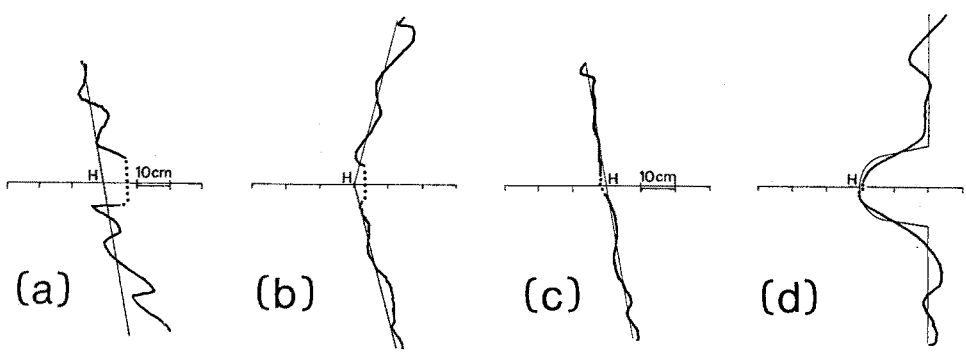

Figure 3. Results of large field simulations. Same legend as in Fig.2, except (d) where the sphere of radius 10 cm is located in front of a planar background.

3. Discussion.

Our simulations show that stereo-correspondence can be established by using the velocity field. However, except for particular 3-D motions, the SCOF is generally not sufficient for reconstructing the depth-map of a whole scene. Rather, it is likely to cooperate with intensity-based stereopsis and motion parallax, essentially because it can be applied to non-rigid environments and to 3-D motions that are useless for most structure-from-motion algorithms (translation in depth, rotation around the sagittal axis).

Several neurophysiological studies [3,9,11] have demonstrated the existence of neurons coding specifically the binocular disparity of retinal motion in the visual pathway of the cat and monkey. In addition, two psychophysical studies [1,5], suggest that motion disparity can be used by the visual system for the perception of *3-D motion*. We have performed some experiments [2] to find out whether *3-D structure* could also be perceived on the basis of the disparity between left and right optical flows. Our results suggest that the SCOF is not used as a depth-cue by the human visual system in the absence of intensity-based stereopsis (the left and right images were stereo-correlated in motion, but not in intensity). We advance the hypothesis that motion disparity can be used by the visual system as a cue for 3-D motion, but not for the fine analysis of 3-D structure.

REFERENCES

(1) Beverley K.I., Regan D., 1973: Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. J. Physiol., 235, 17-29.

(2) Cornilleau-Pérès V., Droulez J., 1990: Motion disparity in the absence of position disparity. A binocular visual cue for depth perception? Perception, 18, 535.

(3) Cynader M., Regan D., 1978: Neurons in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. J. Physiol., 274, 549-569.

(4) Jenkin M.R.M., 1984: The stereopsis of time-varying imagery. Technical Report RBCV-TR-84-3. University of Toronto.

(5) Lee D.N., 1970: Binocular stereopsis without spatial disparity. Percept. Psychophys., 9 (2B), 216-218.

(6) Mitiche A., 1984: On combining stereopsis and kineopsis for space perception. Proc. First Conf. on AI Applications, 156-160.

(7) Mitiche A., 1988: Three-dimensional space from optical flow correspondence. Computer Vision, Graphics and Image Processing, 42, 306-317.

(8) Richards W., 1983: Structure from stereo and motion. M.I.T., A.I. Memo n°731.

(9) Toyama K., Kozasa T., 1982: Responses of Clare-Bishop neurons to three-dimensional movement of a light stimulus. Vision Res., 22, 571-574

(10) Waxman A.M., Duncan J.H., 1985: Binocular image flows: steps toward stereo-motion fusion. Univ. of Maryland, Computer Vision Laboratory, Report CAR-TR-119. May 1985.

(11) Zeki S.M., 1974: Cells responding to changing image size and disparity in the cortex of the rhesus monkey. J. Physiol., 242, 827-841.