

Relative Depth from Vergence Micromovements

Antônio Francisco

Computational Vision and Active Perception Laboratory (CVAP) *
Royal Institute of Technology (KTH), Stockholm, Sweden

Abstract

Relative depth information can be obtained using extremely fine vergence movements (vergence micromovements) about the fixation point with almost no computation and without knowledge of camera parameters. The vergence micromovements approach uses a continuous vergence angle control with simultaneous computation of the local correspondence response of elements with the same relative position in the left and right images. After a complete micromovement cycle a dense relative depth map of the object on the field of view is computed. The relative depth information is stable with respect to the angle of gaze for initial fixation point slightly far from the mid point of the interocular line. Experimental results from physiology and psychophysics suggest that the approach is biologically plausible.

1 Introduction

Relative depth information can be obtained with almost no computation from a sequence of image pairs acquired during extremely fine vergence movements (vergence micromovements). The following procedure is done for every vergence angle step: for each image-point on the left image plane the “correspondence response” of this point and the image-point at the same relative position on the right image plane is stored. After a set of vergence micromovements (micromovement cycle) each image-point has an associated correspondence response signal. Using the correspondence response signals, the relative depth of the 3D points inside the field of view where the correspondence response reach the highest level are computed.

In the present work, for each image-point, instead

of storing the complete correspondence response signal (one value for every vergence angle step) we store only the vergence angle step relative to the highest correspondence response. After completing a micromovement cycle we have a single vergence angle step for each image-point that did not show miscorrespondence (more than one high correspondence response peak). The matrix of these vergence angle steps is the relative depth map of the object in the field of view swept during that micromovement cycle.

Note that there is no need to know the camera geometry, therefore the problem of finding parameters of camera calibration solving a set of non linear equations (registration problem) is avoided. The relative depth map is stable concerning different angles of gaze for a initial fixation point slightly far from the mid point of the interocular line (baseline).

The micromovements approach [2] is functionally different from previous works in the stereo vision area in the sense that we use vergence micromovements to nullify the disparity between the left and right visual stimulus at the same image locus.

2 The stereo vision system

The two-camera model with non zero angle of gaze ν assumes that the principal optical axes intersect at the fixation point and that both X -axes of the two cameras and the fixation point lie on the same epipolar plane (see figure 1). With this assumption, any camera torsion (about the axis connecting the lens center and the image plane center) may be considered to be zero. This model was introduced in [9] for the study of qualitative relative depth from binocular disparities.

In this coordinate system the fixation point is the origin, the epipolar plane of the fixation point is the $X - Z$ plane and the line perpendicular to this plane and passing through the origin is the Y -axis. On the epipolar plane ($X - Z$ plane) the optical axes intersect at the fixation point (origin) with an angle of vergence

*Address: NADA, KTH, S-100 44 Stockholm, Sweden

Email: kico@bion.kth.se, kico@dpi.inpe.br

The author is a visiting researcher from the National Institute of Space Research (INPE), Brazil.

2μ . Let the Z -axis pass through the angle bisector of 2μ and the X -axis be perpendicular to the Z -axis in the $X - Z$ plane. For any point $P = (x, y, z)^T$ in the world coordinate system we define α as the angle of tilt of this point (the angle of its projection on the $X - Z$ plane: $\alpha = \arctan(\frac{z}{x})$). Let (x_l, y_l) and (x_r, y_r) be the Cartesian coordinates of the projection of P on the left and right images respectively.

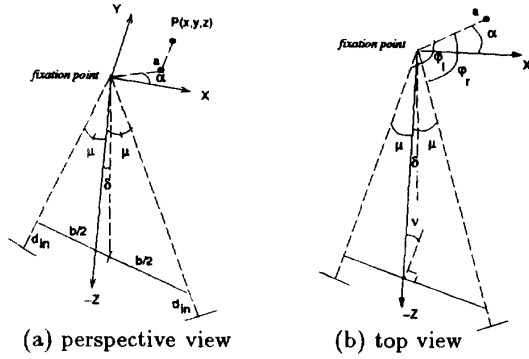


Figure 1: Stereo camera geometry

A case previously studied in [3] is the stereo vision system with symmetric fixation in the visual plane ($\nu = 0$ in figure 1). According [3], we can define an "intersection" point in space where the left and right visual rays coming from image-points at the same relative position on the left and right images (conjugate image-points: $x_l = x_r$ and $y_l = y_r$) almost intersect, i.e. have the same values of z and x coordinates and different values of y coordinate. Let us denote the value of y coordinate from the left visual ray equation as y_{lv} and from the right visual ray equation as y_{rv} . The effect of the difference between this "intersection" point and the ideal intersection point (same x , y and z coordinates) is negligible for object distances greater than ten baselines [3], i.e. $y_{lv} \approx y_{rv}$.

Note that the concept of "intersection point" has been introduced only to visualize how the correspondence response from conjugate image-points can be used to acquire depth information. As we are analyzing the response of a correspondence operator whose inputs are the neighborhood of the two conjugate image-points, we can expect to have the highest response when the "intersection point" is over the surface of an object since both conjugate points should be acquiring the same image.

Let us extend the concept of "intersection point" to all left and right conjugate image-points of our image sensors devices. The shape of the surface formed by

the "intersection points" relative to all these conjugate points, henceforth called "intersections surface", is shown in figure 2.a. Figure 2.b shows the set of intersections surfaces for different vergence angles. We can imagine this set of surfaces as a volume in space, henceforth called "intersections volume".

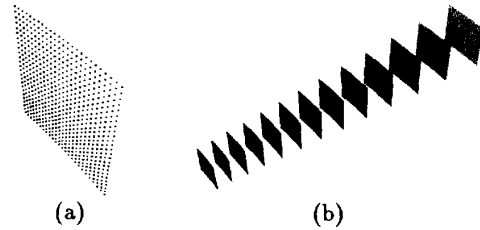


Figure 2: (a) Intersections surface for a given vergence angle step, (b) Intersections volume

The study described in [3] is now extended to the more general case of a system with non zero angle of gaze. For the angles φ_l and φ_r defined in figure 1.b and from [9] we can deduce the following equation for any "intersection point" (same values of z and x coordinates and different values of y coordinate) in space:

$$\frac{x_r}{y_r} \frac{x_l}{y_l} = \frac{\sin(\varphi_r)}{\sin(\varphi_l)} \frac{y_{vl}}{y_{vr}} \quad (1)$$

Assuming that the "intersection point" is close enough to the ideal intersection point ($y_{vl} \approx y_{vr}$) we can apply (1) to estimate the shape and direction of the "intersections surface" containing the entire set of intersection points derived from all conjugate image-points. Therefore, since for each conjugate image-points $x_l = x_r$ and $y_l = y_r$, from (1) we get $\sin(\varphi_r) \approx \sin(\varphi_l)$.

Additionally, from figure 1.b, we get:

$$\begin{aligned} \varphi_r &= \alpha + 90^\circ - \mu \\ \varphi_l &= \alpha + 90^\circ + \mu \end{aligned}$$

Substituting the above values of φ_l and φ_r in $\sin(\varphi_r) \approx \sin(\varphi_l)$ and expanding, we get $\sin(\alpha) \approx -\sin(\alpha)$ (since $\mu \neq 0$). This expression is valid for small values of α . The conclusion that α (tilt) must be nearly zero for any "intersection point" in space is that its z coordinate must also be nearly zero since $\alpha = \arctan(\frac{z}{x})$ (see figure 1.b). In other words, the relative depth of this point with respect to the origin (fixation point) is nearly zero. Therefore the intersections surface containing the "intersections" from all conjugate points ($x_l = x_r$ and $y_l = y_r$) is nearly the $X - Y$ plane.

In order to have a relative depth independent of the angle of gaze, the angle δ between the Z -axis and the line connecting the fixation point and the mid point of the baseline should be zero. We will demonstrate that δ is near zero under certain conditions. The analysis is done considering the object distance as a multiple of the baseline ($d_{obj} = k_b b$). Let us develop the expression for δ using the sine and cosine theorems and the geometry shown in figure 1.b for a positive angle of gaze ν ,

$$\delta = \arcsin\left(\frac{(1/2) \cos \nu}{\sqrt{k_b^2 + (1/2)^2 - k_b \sin \nu}}\right) - \arcsin\left(\frac{(1/2) \cos \nu}{\sqrt{k_b^2 + (1/2)^2 + k_b \sin \nu}}\right) \quad (2)$$

Thus, for an object distance d_{obj} great enough so that $k_b \gg \sin \nu$, we obtain the relative depth map of the object in the field of view independent of the angle of gaze (since δ is very small). Values of k_b greater than ten satisfies the previous condition and minimizes the effect of the difference between the "intersection point" and the ideal intersection [3].

We have done the analysis of the shape and direction of the intersections surface for a single value of vergence angle 2μ . The same analysis can be expanded to continuous change on the vergence angle.

3 Continuous vergence angle control

The concepts of "intersection" points and intersections surface previously introduced for a single vergence angle 2μ is extended for the multiple vergence angle case. The main idea is that a continuous vergence angle change is done about the initial fixation point producing what we call a complete vergence micromovement cycle. Therefore the analysis now is conducted for a number of vergence angles $2\mu_i$ about the *fixation point in the visual plane*, described by the following expression:

$$\mu_i = \mu + \varepsilon_i \quad (3)$$

where ε_i is a small camera vergence angle step (positive or negative) that describes the micromovement about the *fixation point*. The set of μ_i about a given fixation point determines a complete vergence *micromovement cycle*.

The resulting set of intersections surfaces, one for each vergence angle, follows the main scheme shown in figure 2.b for a stereo vision system with symmetric fixation (zero angle of gaze). In the case of non

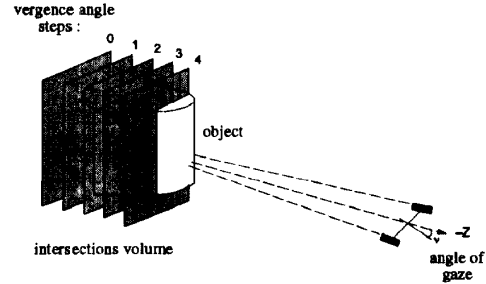


Figure 3: Object being swept

zero angle of gaze these almost planar surfaces are parallel to the $X - Y$ plane therefore they move in the Z -axis direction as the vergence micromovement is performed. As we mentioned before, we can imagine this set of planar surfaces as a volume in space where the intersections points will change their positions during a complete micromovement cycle (figure 3). Each intersection point in space for a given vergence angle moves to another position on the adjacent intersections surface of the next vergence angle. In other words, every intersection point moves in space along an imaginary line almost on the same direction as the Z -axis but centered on the intersection position determined for the initial fixation point.

The resulting effect of this movement is that any object in space inside the intersections volume will be swept in depth by the intersections during a complete micromovement cycle (figure 3). According to the micromovement approach [2], any object inside this volume can have its depth measurements determined by the response of a *local correspondence operator* to the *continuous vergence angle control*. The inputs to this operator are the left and right image neighborhoods centered on the conjugate image-points. It is acceptable to use a local correspondence operator since we assumed that d_{obj} is greater than ten baselines (see previous section).

In order to illustrate the dynamic process behind the micromovement approach, let us analyze the correspondence response for different vergence angles. Figure 4 shows the correspondence response relative to the same conjugate image-points for different vergence angles. As mentioned previously the highest correspondence response occurs when the intersection point pass over the surface of the object, since both image neighborhoods around the conjugate image-points should be of the same part of the object. Therefore, the relative depth for that conjugate image-points is the vergence angle where the highest correspondence peak is detected.

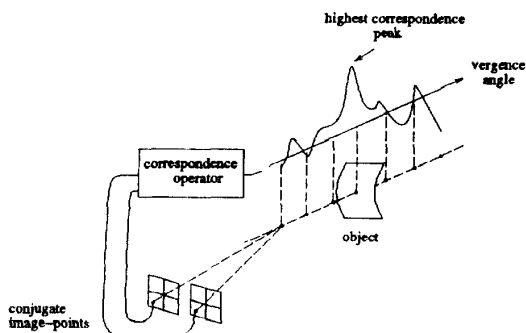


Figure 4: Correspondence response

It is important to emphasize that the relative depth information between adjacent intersections surfaces is maintained since the vergence micromovements shift the surfaces on the Z -axis direction (depth direction). The resulting effect after a complete micromovement cycle is a dense relative depth map of the object in the field of view inside the intersections volume.

4 Biological support

The following discussion of eye-movements according to physiological and psychophysical experiments is offered as a working hypothesis, useful for the understanding the role of the micromovements on relative depth perception. Physiological results [5, 1] show that the human eye performs fine movements during the process of fixation on a single point, which are collectively called *physiological nystagmus*. Physiological nystagmus is composed of three different kinds of movements: (1) high-frequency tremor, (2) slow drifts, and (3) rapid *binocular flicks*. The drift and flick movements occur in opposing directions and produce convergence/divergence waves of the eyes on a similar way as the micromovements studied in previous sections.

Assuming the vergence micromovements mechanism as the basis of the relative depth perception, it is easy to understand the phenomena of *stereoacuity* (depth or stereoscopic acuity, stereopsis). As well described in [5, 8], it is almost incredible that most observers under normal conditions can discriminate a difference in depth (i.e. relative depth) corresponding to an *angular disparity* (interocular disparity) of about 10 arc sec. The best values reported in the literature have been obtained by the apparatus called the *Howard-Dolman apparatus*, devised by Howard in 1919. The best observers achieve a 75% discrimina-

tion level close to 2 arc-seconds in that experiment. The most incredible fact is that this disparity value is much smaller than the distance between the cones' centers at the *central part of the fovea* (≈ 22 arc sec).

We suggest that the high sensitivity to slight disparity can be explained by the correlation between relative depth perception and the vergence micromovements and not by the capacity of the human visual system to spatially detect disparity on the retinas. Therefore the idea of an *angular disparity* that can be detected *spatially* by the visual system is substituted by a *local approach* where the human visual system determines the relative depth values by the highest peak of correspondence response during a complete *micromovement cycle*. The highest peak of correspondence occurs when there is no spatial disparity between the left and right stimulus of elements with the same relative position on both retinas, i.e., when the spatial disparity is cancelled for a given vergence angle.

Another phenomenon that can be explained by the present approach is known in the literature [5] as *Panum's fusional area*: the range of interocular disparities within which objects viewed with both eyes on corresponding retinal regions appear single. This area is such that *fusion* occurs, only one dot is seen, when two points that are perceived in different eyes fall closer together in the combined view. Note that these two points can be seen through an uncrossed (left and right optic axis do not cross) or crossed disparity. The classical static limits for Panum's area, the mean crossed to uncrossed range of horizontal disparities, is reported as being 14 arc min. The experiments described in [4] support the existence of *binocular fusion* as a *unique* category of sensory performance, disconfirming several non fusional explanations of single vision. While the range of binocular disparities allowing fusion (Panum's fusional horizontal diameter) is typically in the region of 14 arc min, stereoscopic relative depth can be perceived from a disparity 500 times smaller.

In the present approach, the phenomena of binocular fusion and stereoscopic relative depth are assumed to be supported by the mechanism of vergence eye micromovements about a fixation point. In this way, the fusion area dimension is determined by the range of a complete *micromovement cycle*. It is important to point out that the *classical* value of the Panum's fusional *horizontal radius* (average of the crossed and uncrossed disparities), 7.0 arc min, coincides with the micromovement range value described in [1]. Note that the Panum's fusional horizontal radius must be compared to the total range of a *monocular* micromove-

ment reported in [1] to be coherent to both definitions. In the present analysis the *vertical* fusion radius is not considered since this radius follows the monocular spatial resolution limit of the retina [7]. As a conclusion, the “real” binocular fusion is assumed to occur only between cells adjacent on the horizontal axis of the retinas, and that binocular vertical fusion is a result of the monocular fusion mechanism.

5 Experiments

In order to validate experimentally the relative depth strategy from vergence micromovements, the KTH head-eye system was used [6]. This stereo vision system has several step motors to control the vergence angle of both cameras, the angle of gaze, zoom, focus, size of the baseline and other positional motors. Both left and right frame grabbers are directly connected to a transputer board localized on the VME bus of a SUN-station. The set of experiments was done on the OCCAM environment of the transputer network in order to improve the time performance. The best time performance reached up to now was around 0.2 seconds to compute the correspondence response for each pair of images of 200x200 pixels. The correspondence operator used is a simple difference operator. For each conjugate image-points, we have an operator window around it. We take the sum of the square of the difference between the left and right images for each pixel inside this operator window.

The aim of these experiments is to demonstrate that relative depth information stable with respect to the angle of gaze can be obtained without to much computation and without the knowledge of the camera parameters. The initial set up is done for zero angle of gaze by choosing a value for the focal length (zoom) and adjusting the focus and the initial fixation point over the central part of the object surface. Every time we change the angle of gaze the system is vergence adjusted in order to point the initial fixation point and focus (rarely zoom) adjusted in order to keep the left and right image with similar height. This is done because a non zero angle of gaze puts one camera farther from the object than another consequently the image becomes bigger in one camera than another.

For each angle of gaze, a program makes the sweeping of the object by changing the vergence angle between the initial vergence angle and a final vergence angle determined by the number of vergence steps specified. For each vergence step the program acquires

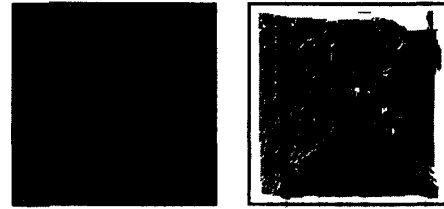


Figure 5: Newspaper. (a) Object image, (b) Relative depth

the left and right images and computes the correspondence response. The maximum peak of correspondence for every conjugate image-points acquired up till this time is stored on the relative depth map.

The following parameters are common in both experiments: $b = 200$ mm, $d_{obj} = 12b$, work window of 200x200 pixels, operator size of 21x21 pixels and the camera vergence angle resolution of 26 arc sec.

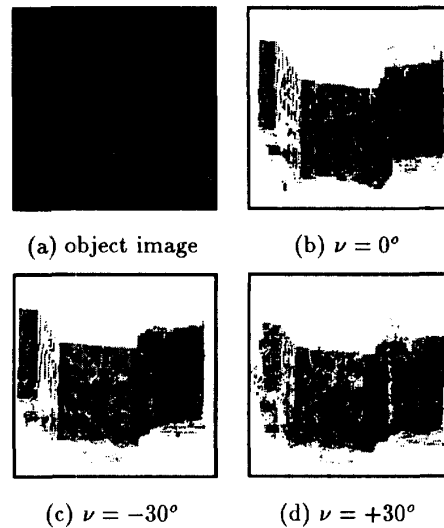


Figure 6: Box-book. (a) Object image; (b), (c), (d) Relative depth for different angles of gaze

The first experiment uses a newspaper as the object under symmetric fixation (figure 5). This experiment was done to demonstrate that the vergence micromovement approach works properly when the angle of gaze is equal to zero.

Our second experiment was done using as the object one box (left side of the object) in front of an inclined book, with respect to the head-eye system. The box and book are over a platform. In figure 6.a the object image acquired for every highest correspondence

peak computed is shown. It is important to point out that the image shown is not the simple grey image acquired from the left or right camera at a given vergence angle, but it is the grey level acquired for each pixel at the vergence angle in which that conjugate pixel computed the highest correspondence response. Therefore the image shown in figure 6.a was composed during the micromovement cycle. The other items of figure 6 show the relative depth acquired after the micromovement cycle.

An analysis of the above experiment shows that the relative depth acquired is stable with respect to the angle of gaze. I.e., the dense relative depth map acquired as a direct application of the vergence micromovement approach gives a robust primal representation of the shape of the object being visualized without the knowledge of the camera parameters. The only constraint is that the fixation point at the beginning of the micromovement cycle and the X-axes of both cameras must be at the same plane.

6 Conclusion

The *vergence micromovements* approach permits to overcome the physical limitation of the photo-receptor dimension on the depth perception. Using vergence micromovements the intensity over each photo-receptor varies so little for adjacent vergence angles that a best match between the left and right images can be reached after a complete micromovement cycle. Other advantages of the present approach are due to its local characteristic for searching the highest correspondence response for each conjugate image-points. First, the use of the same set of photo-receptors during the search for the highest peak overcomes the existing differences between photo-receptors of any image device. Therefore we can expect to have better experimental results than if using only two images and searching for matching. Second, the method can be implemented in a modular hardware scheme suitable for using in real time active vision systems. A good compromise between performance and implementation costs can be obtained as shown in figure 7, where the correspondence operator is implemented as an analog device and the correspondence images are acquired via a frame grabber.

The highlight of this new approach is the vergence micromovements as a mechanism to nullify the disparity between the left and right visual stimulus at the same image locus.

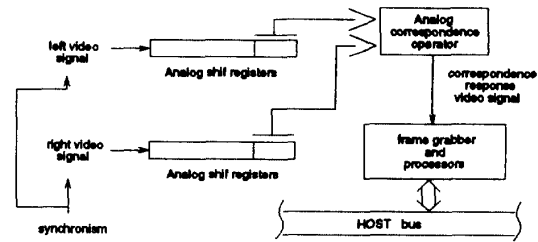


Figure 7: Analog implementation

Acknowledgements

The support from the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged. I would like to thank Prof. Ruzena Bajcsy and Prof. Jan-Olof Eklundh for the support to the development of this work as well as Kourosh Pahlavan and Tomas Uhlin for valuable help when using the KTH head-eye system.

References

- [1] R. W. Ditchburn. Eye-movements in relation to retinal action. *Optica Acta*, 1(4):171-176, 1955.
- [2] A. Francisco. The role of vergence micromovements on depth perception. Technical Report MS-CIS-91-37, GRASP LAB, CIS, University of Pennsylvania, Philadelphia, PA, USA, 1991.
- [3] A. Francisco. Vergence micromovements and depth perception. In *British Machine Vision Conference*, pages 367-376, Leeds, Great Britain, September 1992.
- [4] T. Heckmann and C. M. Schor. Panum's fusional area estimated with a criterion-free technique. *Perception & Psychophysics*, 45(4):297-306, 1989.
- [5] J. W. Kling and L. A. Riggs. *Experimental psychology*. Holt, Rinehart and Winston, Inc., 1971.
- [6] K. Pahlavan and J.O. Eklundh. A head-eye system - analysis and design. In *Computer Vision, Graphics, and Image Processing: Image Understanding*, pages 41-56, July 1992.
- [7] C. Schor, I. Wood, and J. Ogawa. Binocular sensory fusion is limited by spatial resolution. *Vision Res.*, 24(7):661-665, 1984.
- [8] C. M. Schor and K. J. Giuffreda. *Vergence eye movements: basic and clinical aspects*. Butterworth, 1983.
- [9] D. Weinshall. Qualitative depth from vertical and horizontal binocular disparities, in agreement with psychophysical evidence. *IEEE Computer Vision and Pattern Recognition*, pages 159-164, 1988.