May 1991

The Role of Vergence Micromovements on Depth Perception

Antônio Francisco Júnior
University of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-91-37.

The Role of Vergence Micromovements on Depth Perception

Abstract

A new approach in stereo vision is proposed which recovers 3D depth information using *continuous vergence angle control* with simultaneous local correspondence response. This technique relates elements with the *same* relative position in the left and right images for a continuous sequence of vergence angles. the approach considers the extremely fine vergence movements about a given fixation point within the depth of field boundaries. It allows the recovery of 3D depth information given the knowledge of the system's geometry and a sequence of pairs $[\alpha_i, C_i]$, where α_i is the i^{th} vergence angle and C_i is the i^{th} matrix of correspondence responses. The approach has several advantages over the current ones. First, due to its local operation characteristics, the resulting algorithms can be implemented in a modular hardware scheme. Second, unlike currently used algorithms, there is no need to compute depth from disparity values; at the cost of the acquisition of a sequence of images during the micromovements. The approach also greatly reduces the errors in stereo due to the sensor quantization. Last, and most important of all, the approach is supported by experimental results from physiology and psychophysics. Physiological results show that the human eye performs fine movements during the process of fixation on a single point, which are collectively called *physiological nystagmus*. One such movement, called *binocular flicks*, happens in opposing directions and produces convergence/divergence of the eyes. These are the micromovements that we suppose are the basis for depth perception. Therefore, the approach proposes a functional correlation between these vergence micromovements, depth perception, stereo acuity and stereo fusion.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-91-37.

The Role Of Vergence Micromovements On Depth Perception

MS-CIS-91-37
GRASP LAB 262

Antônio Francisco Júnior

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

May 1991

The Role of Vergence Micromovements on Depth Perception

Antônio Francisco Júnior

General Robotics and Active Sensory Perception Laboratory

School of Engineering and Applied Science

University of Pennsylvania, Philadelphia, PA 19104

Abstract

A new approach in stereo vision is proposed which recovers 3D depth information using *continuous vergence angle control* with simultaneous local correspondence response. This technique relates elements with the *same* relative position in the left and right images for a continuous sequence of vergence angles. The approach considers the extremely fine vergence movements about a given fixation point within the depth of field boundaries. It allows the recovery of 3D depth information given the knowledge of the system's geometry and a sequence of pairs $[\alpha_i, \mathbf{C}_i]$, where α_i is the i^{th} vergence angle and $\mathbf{C}_i$ is the i^{th} matrix of correspondence responses. The approach has several advantages over the current ones. First, due to its local operation characteristics, the resulting algorithms can be implemented in a modular hardware scheme. Second, unlike currently used algorithms, there is no need to compute depth from disparity values; at the cost of the acquisition of a sequence of images during the micromovements. The approach also greatly reduces the errors in stereo due to the sensor quantization. Last, and most important of all, the approach is supported by experimental results from physiology and psychophysics. Physiological results show that the human eye performs fine movements during the process of fixation on a single point, which are collectively called *physiological nystagmus*. One such movement, called *binocular flicks*, happens in opposing directions and produces convergence/divergence of the eyes. These are the micromovements that we suppose are the basis for depth perception. Therefore, the approach proposes a functional correlation between these vergence micromovements, depth perception, stereo acuity and stereo fusion.

1 Introduction

The perception of the 3D-distance, depth, of objects using stereo images have been studied by many researchers for a long time. Initially, the propose was to recover the object depth using parallel cameras, since this appeared as the easiest way to acquire stereo images without mechanical moving parts. Also, using parallel cameras geometrically generate parallel epipolar lines that are simpler to treat when implementing a search algorithm for matches. Although the parallel cameras system is simple in some aspects, it presents errors in depth due to quantization that are not negligible. These errors are due the finite area of the photoreceptor on the CCD device [20].

A new approach was introduced using vergence camera systems ([12]). In this way, using the controls over vergence angle and focus, the corresponding features are first identified between the left and right images and then their depth is estimated. This integration of position control, image acquisition and depth processing inaugurated a new modality of vision system named "active vision" [3, 4].

While intending to overcome the traditionally most difficult problem in stereo vision, some new algorithms have been introduced to find the corresponding features. There has already been some research in this area [5, 21, 10, 1].

In many of these methods, the cameras movements are relevant as a fundamental tool to acquire and process stereo matching. As many researchers pointed out (e.g. [15]), the role of eye movements is relevant on the recovery of stereo depth.

This paper analyzes the extremely fine movements (micromovements) of the cameras about the fixation point. It assumes that these movements are synchronized between the two cameras. Therefore, the new approach here is the role of these micromovements on the depth perception.

This kind of movements differs from the other ones studied until today, since the vergence,

translation and rotation movements were used functionally to fixate the cameras on a new fixation point. On this new fixation point, the previous methods (using particular techniques as multiresolution) compute some depth-map or depth directly by the acquisition of the left and right images at this point, i.e., using these images and some stored information (estimation) about the depth-map they are able to infer the current depth at the images' points matched correctly.

It is important to emphasize that the strategies used in all the previous methods have the search space for correspondence matches over epipolar lines. This means that the depth is calculated using the disparity information between the left-right matched points and the geometry of the camera system.

The present method uses neither epipolar lines, nor disparities, to calculate the depth of any 3D point. The depth is determined by the geometry of the camera system (mainly, the vergence angle position) and by the relative position of the pixel with respect to the image plane. The procedure can be simply described as follow: micromovements of two cameras occur about the fixation point. For each left-image point, on the left image plane, is stored the vergence angle and the "correspondence response" of this point and the right-image point at the same relative position on the right image plane. For each left-point, using these "correspondence response" signals and the camera geometry, the depth of the 3D points where the correspondence response reach a desired high level is calculated.

Therefore, the present approach is functionally different from the previous one in the sense that the depth is calculated locally for each point (without searching epipolar lines) with the necessity of acquiring a sequence of images during the micromovements.

It is not the intention of the present work to analyze either the role of the multiple scale (multi channels) on the depth processing, or the best "correspondence operator" to be used. Therefore, the problems of the correspondence will be explored in future work. The objective here is to clarify how to calculate depth using micromovements. Some psychological and

physiological findings support the present approach on the human visual system.

2 The stereo vision system

Each convex lens of the right and left camera is considered *ideal*, in the following sense: if it is illuminated with a collimated beam of light parallel to the optical axis, the beam cross the optical axis at the same point named focal length (**f**) of the lens. The *aberrations* are not discussed in this work. Additionally, the *thin lens* approximation to the camera lenses is assumed. Therefore, a single plane can be used to define the position of the lens along the optical axis. In this model, when an object is at a distance $\mathbf{d_{out}}$ (the *object distance*) from the principal plane (see Figure 1) its image appears, with inverted direction, at the distance $\mathbf{d_{in}}$ (the *image distance*) from this same plane. The relationship of these two distances and the focal length of the lens is given by the *Gaussian Lens Equation*:

$$\frac{1}{f} = \frac{1}{d_{in}} + \frac{1}{d_{out}} \tag{1}$$

With respect to the camera platform, a symmetric fixation in the visual plane is assumed. Therefore the vergence angle of the two cameras have the same value α and the point being fixated is in the visual (horizontal) plane of the cameras. With this assumption, any camera torsion (about the axis connecting the lens center and the image plane center) may be considered to be zero. According the symmetric fixation model (figure 2) associated with each image plane there is a coordinate system with its origin at the image plane center. The lenses centers are separated by a baseline **b**. These coordinate systems define the left projection (x_{pl}, y_{pl}) and right projection (x_{pr}, y_{pr}) of an object point in space. To identify the 3D position $(X_o, Y_o, Z_o)^T$ of an object point $\vec{P}_o$ in space a global coordinate system *xyz* is used, as shown in Figure 2.

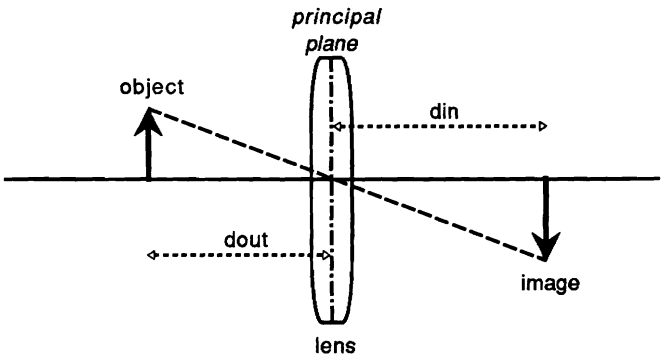


Figure 1: Thin Lens Geometry

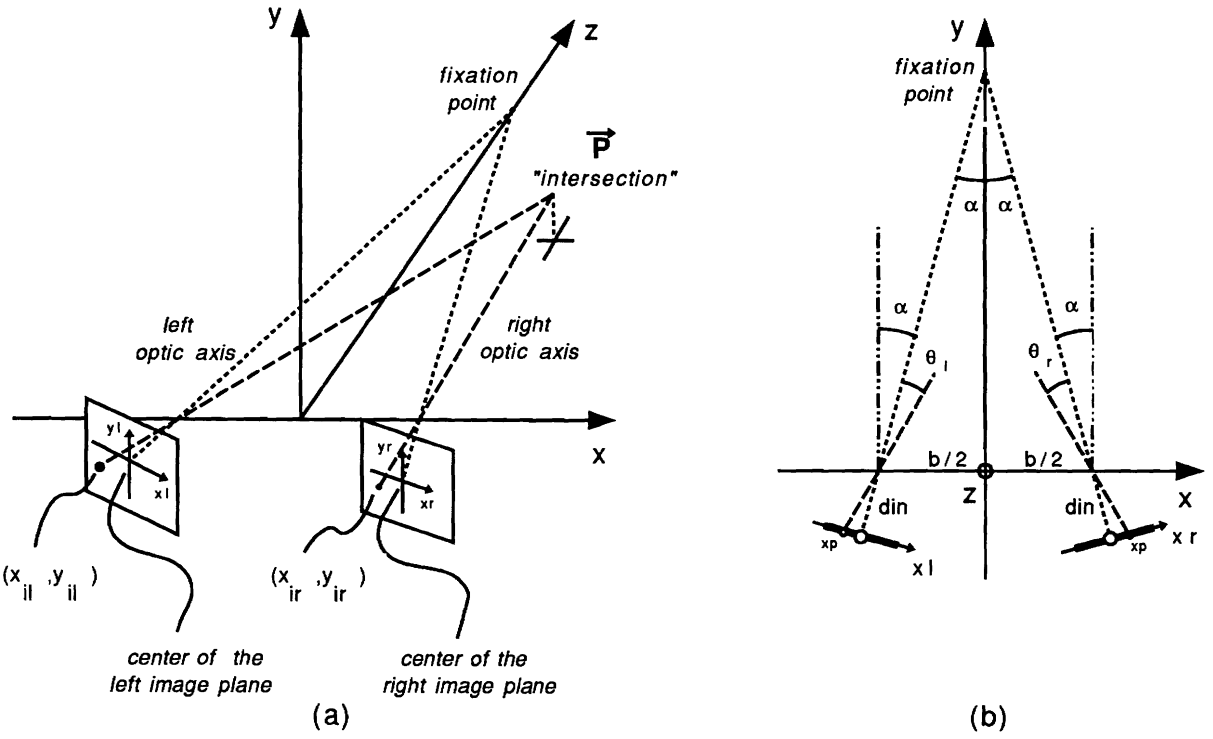


Figure 2: Stereo cameras geometry: (a) Perspective view, (b) Top view

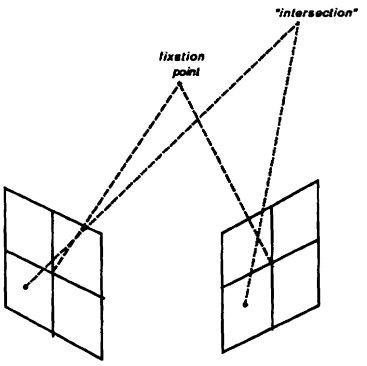


Figure 3: Geometric image planes correspondence

Based on the above coordinate systems, a new approach is presented to recover 3D depth information using **continuous vergence angle control** with simultaneous local correspondence response. This technique relates elements (image plane points) with the *same* horizontal and vertical distance from the center of the left and right image planes for a continuous sequence of vergence angles. These elements have zero binocular disparity on the two image planes (figure 3).

It is important to point out that the two-cameras system analysis is conducted in accordance to the classical human eye approach. To form this relation some concepts that have been used in the physiology and psychophysics sciences are introduced. One of these concepts, the **horopter**, defines the set of points in space for which the binocular disparity is zero [18]. The **point horopter** is the locus of zero disparities for the point stimulus where *both* horizontal and vertical disparities are zero. Accordingly [18], there has been a considerable amount of confusion in the literature caused by the laxity in defining the horopter.

It must be emphasized that the ideal point horopter is that one with zero horizontal and vertical disparities for the symmetric fixation in the visual plane, with any position torsion and optical aberrations assumed to be absent. For the human eye, this ideal horopter has a shape shown in Figure 4, although in certain restricted circunstances it may become a two-

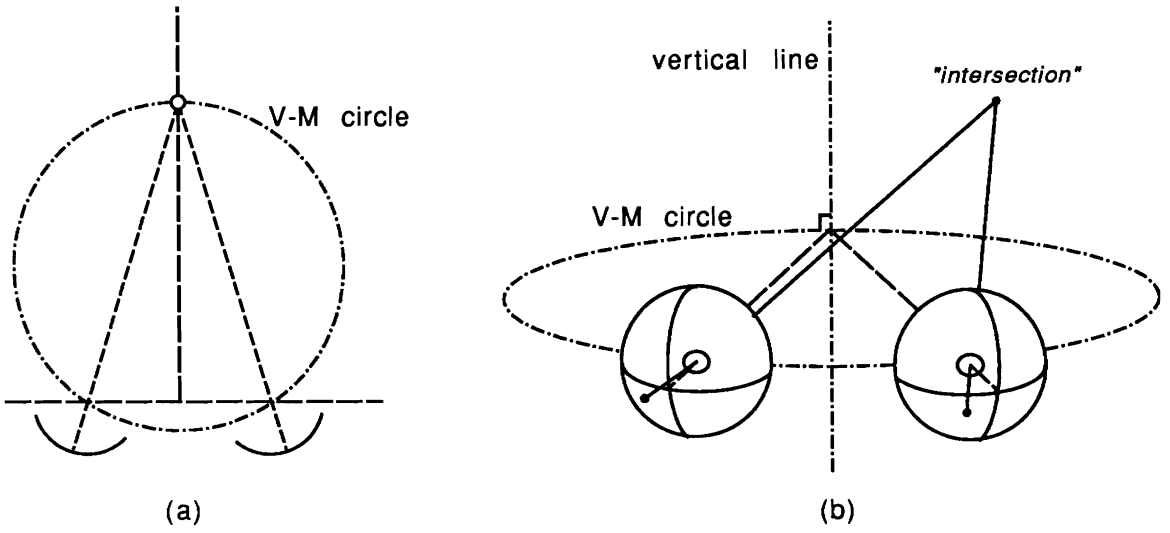


Figure 4: Shape of the point horopter for the symmetric fixation in the visual plane: (a) Top view, (b) Perspective view

dimensional surface [18]. Note that the component formed by the correspondence of the points along the horizontal midline through the image plane center (fovea in the human case) of the symmetric point horopter is designated as the Vieth-Muller (V-M) circle [18].

As described in [18], and shown in Figure 4, any point off the *point horopter* in space (off-axis points) project to the two image planes with horizontal and vertical disparities. With vergence, the points at the distance corresponding to the point horopter would nullify the horizontal disparity. Note that in the ideal case nothing can be done to nullify the vertical disparity produced by off-axis points being necessarily closer to one eye than the other, with a resulting difference in the projection angle in the two eyes.

The present analysis concerning the horopter is based on *zero horizontal* disparity. It is shown that the point horopter could be extended to a surface shape due to some human visual system characteristics such as the enlarging of the receptive field on the periphery and the expected difference between nasal and temporal retina eccentricity [18].

2.1 Computing the horopter

Since the present analysis is concerned with *zero horizontal disparity* it is possible to define a 3D "intersection" point of the left and right optic axes (same x and z and different ys) passing through the same corresponding element $(x_p, y_p)^T$ on the image planes coordinate systems (Appendix A). The coordinate $(x_\iota, y_\iota, z_\iota)^T$ of this "intersection" point is described by the following expressions:

$$x_\iota = \frac{b}{2}\frac{\tan(\alpha+\theta_l)-\tan(\alpha-\theta_l)}{\tan(\alpha+\theta_l)+\tan(\alpha-\theta_l)} \tag{2}$$

$$y_\iota = -\frac{y_p}{d_{in}}\frac{b}{2}\frac{cos^2(\theta_l)}{\sin(\alpha)} \tag{3}$$

$$z_\iota = \frac{b}{\tan(\alpha+\theta_l)+\tan(\alpha-\theta_l)} \tag{4}$$

where,

$$\alpha = \arctan(\frac{b}{2d_{out}}) \tag{5}$$

$$\theta_l = -\arctan(\frac{x_p}{d_{in}}) \tag{6}$$

and from (1),

$$d_{in} = \frac{d_{out}f}{d_{out}-f} \tag{7}$$

The above equation for y_ι is deduced considering the mean of the left and right y coordinates of the left and right optic axes at the "intersection" point. This assumption includes an error in the present analysis. The effect of this error is evaluated doing the opposite procedure of the one described above.

Let us pick a single point $\vec{P_o}$, at the position $(X_o, Y_o, Z_o)^T$ in the 3D space, and analyze the difference between the projections on both image planes. Note that the projections are

determined by the intersection of the right and left optic axes, passing through the left and right optic lens center respectively, with the image planes. Let us designate the intersection of the left optic axis with the left image plane as $(x_{pl}, y_{pl})^T$ and the intersection of the right optic axis with the respective image plane as $(x_{pr}, y_{pr})^T$. It is possible to define the following euclidian *projections deviation* (**dev**) for a given point $\vec{P}_o$ as:

$$dev = \sqrt{(x_{pl} - x_{pr})^2 + (y_{pl} - y_{pr})^2} \tag{8}$$

where (see Appendix A),

$$x_{pl} = -d_{in} \tan(\arctan(\frac{\frac{b}{2} + X_o}{Z_o}) - \alpha) \tag{9}$$

$$y_{pl} = \frac{-Y_o d_{in} \sin(\alpha + \theta_l)}{\cos(\theta_l)} (X_o + \frac{b}{2})^{-1} \tag{10}$$

$$x_{pr} = -d_{in} \tan(-\arctan(\frac{(\frac{b}{2} - X_o)}{Z_o}) + \alpha) \tag{11}$$

$$y_{pr} = \frac{Y_o d_{in} \sin(\alpha + \theta_r)}{\cos(\theta_r)} (X_o - \frac{b}{2})^{-1} \tag{12}$$

and,

$$\theta_l = -\arctan(\frac{x_{pl}}{d_{in}}) \tag{13}$$

$$\theta_r = \arctan(\frac{x_{pr}}{d_{in}}) \tag{14}$$

□

The above expressions are valid for any point $(X_o, Y_o, Z_o)^T$ on the 3D space. It is important to emphasize that, as the present work is concerned with small regions about the fixation point, the choice of the object point should be restricted to a region near the fixation point. In order to simplify the choice for a point $\vec{P}_o$, at the position $(X_o, Y_o, Z_o)^T$ in the 3D space, a particular

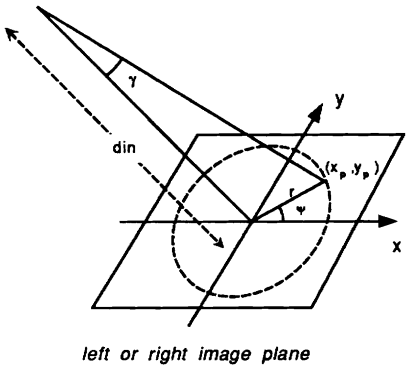


Figure 5: Polar coordinate system of the image plane

point $\vec{P}_o$ in space is used. The point considered is at the position $(x_\iota, y_\iota, z_\iota)^T$ (equations 2 through 4) generated by the "intersection" of the left and right optic axes passing through some desired corresponding position $(x_p, y_p)^T$ with a given vergence angle. In order to facilitate the *dev* analysis:

- The object distance is considered a multiple of the baseline,

$$d_{obj} = k_b b \tag{15}$$

- The image planes are mapped by a polar coordinate system (figure 5), then

$$x_p = d_{in} \tan(\gamma) \cos(\varphi) \tag{16}$$

$$y_p = d_{in} \tan(\gamma) \sin(\varphi) \tag{17}$$

- The *dev* is plotted normalized with respect to the distance between centers of adjacents photoreceptor elements (**dce**).

□

			dce (0.001 mm)			
	f (mm)	b (mm)	γ = -1 deg	γ = 0 deg	γ = 1 deg	γ_{max} with high resolution
GRASP cameras	17.5 to 105	128	30.0	30.0	30.0	The resolution is γ independent.
HUMAN eyes	17	65	5.40	1.81	5.40	0.5 deg (fovea)

Table 1: General data for the GRASP cameras and the human visual system. The *dce* values for the human visual system are rough data.

Having deduced all needed equations for the *dev* analysis, it is time to introduce some real data about the human visual system and the GRASP platform system. Concerning the human visual system, at the *central* part of the *fovea* ($\gamma \leq 20$ arc min) the cones are tightly grouped together and the distance between the cones' centers is ≈ 22 arc sec [11, 8], which is equal to:

$$\frac{22}{3600}\frac{\pi\, b}{180} = \frac{22}{3600}\frac{\pi\, 17\, mm}{180} \approx 1.81\ 10^{-3}\ mm$$

The above data and all other data needed for the present analysis are summarized in table 1.

Note that in the human visual system the *distance between centers* can be fit by a quadratic function, using the data of table 1, as shown in Figure 6. This least-square fitting not necessarily corresponds to the reality, serving only to explain how the human eye system can treat the *dev*. Therefore, in the human visual system the distance between centers of adjacent photoreceptor elements is represented by the function $dce(\gamma)$. In the GRASP platform, since this distance is constant independently of γ value, it is represented by $dce(.)$.

It must be emphasized that the previous equations 9 through 12 are deduced for *planar shape* image planes. The same equations are used to give an idea about the human eye visual system (*spheric shape*). This approximation is quite satisfactory since the *dev* analysis is

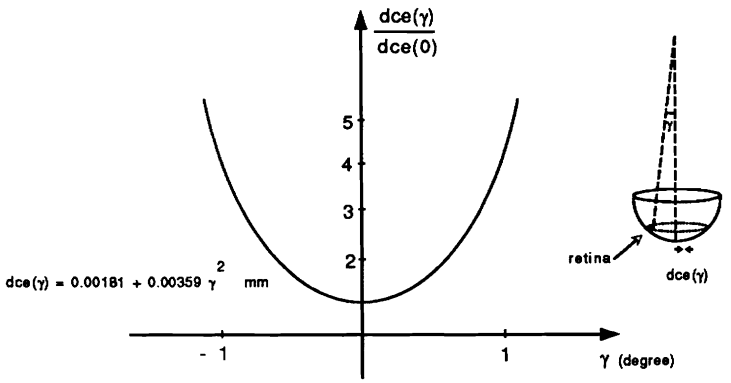


Figure 6: Least-squares quadratic fitting for the distance between centers of adjacent cones in the human visual system

done for photoreceptor elements near the center of the image planes (where γ assumes small values) where the results could be geometrically expected to be similar for the planar and spheric shape image plane.

The procedure to realize the *dev* analysis can be synthesized in the following simulation steps:

- for a given: $\mathbf{k_b}$, γ, φ, $\mathbf{f}$ and $\mathbf{b}$,
- calculate: d_{obj} (equation 15), d_{in} (equation 7), α (equation 5), x_p (equation 16), y_p (equation 17), $(X_o, Y_o, Z_o)^T = (x_\iota, y_\iota, z_\iota)^T$ (equations 2 through 4),
- with the previous value of $(X_o, Y_o, Z_o)^T$, calculate: x_{pl}, y_{pl}, x_{pr}, y_{pr} (equations 9 through 12) and then the *dev* (equation 8).
- plot: The results of the *dev* normalized with respect to the distance between centers of adjacent photoreceptor elements ($dce(\gamma)$).

□

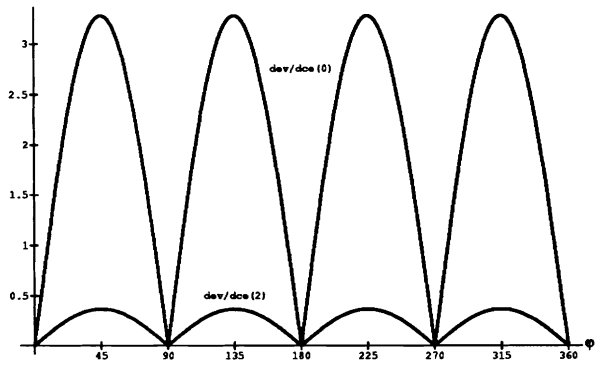


Figure 7: Projections deviation for human visual system as a function of φ ($\gamma = 2^o$, b = 65 mm, f = 17 mm, $d_{obj} = 2b$)

The simulation results shown in Figure 7 imply that the highest *dev* occurs at φ equal 45 degree. Therefore all other simulations are done with the above value of φ. Another conclusion from Figure 7, is that the normalized *dev* is smaller when it is used with a varying $dce(\gamma)$ than with a fixed value of $dce(.)$. This characteristic, implicit in the human visual system, tend to diminish *dev*. Another feature of the human visual system that tend to diminish *dev*, and are not analyzed in the present work, is the expected difference between nasal and temporal retina eccentricity (if nasal is larger than temporal) for every pair of corresponding points. As described in [18] (pp. 213), an object point in the left visual field must project at a *greater* angle from the fovea on the *nasal* retina of the *left* eye than on the *temporal* retina of the *right* eye to hit corresponding receptors. This difference in eccentricity could explain the deviation of the empirical horopter ([18], Figure 7.7) from the V-M circle as well as the necessity to diminish *dev*.

Another good demonstration about the effects of varying $dce(\gamma)$ is shown in Figure 8. In the case of the actual GRASP camera platform is interesting to know how *dev* varies with the object distance. From the analysis of Figure 9 can be seen that *dev* decreases with object distance, therefore the next set of simulations are done with $d_{obj} = 15b$ which ensures a small *dev* for $\varphi = 45$ degree. It must be pointed out that the *dev* has been investigated for its

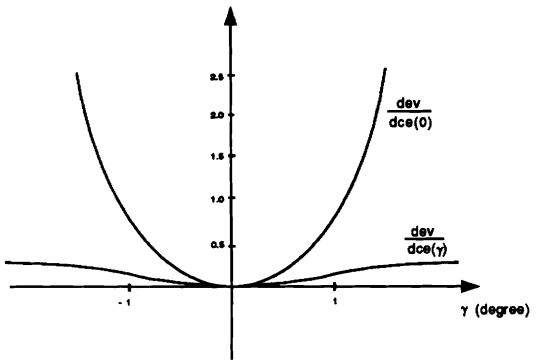


Figure 8: Projections deviation for human visual system as a function of γ ($\varphi = 45^o$, b = 65 mm, f = 17 mm, $d_{obj} = 2b$)

maximum value, in spite of this deviation being zero on the image planes coordinate axes ($\varphi = k$ 90 degree, k = 0,1,..) and very small near these axes for any value of γ, d_{obj}, b and f.

Having shown that the normalized *dev* is very small when d_{obj} is equal to fifteen times the baseline size, the *point horopter* is plotted using equations 2 through 7 for this value of d_{obj}. Almost all the parameters of the above equations can be calculated directly (like d_{in} and α) from the defined values of d_{obj}, b and f. The only two parameters that do not have a defined range are x_p and y_p. The range of these parameters are defined in correspondence with the human visual system, since we are interested in analyzing only the foveal cones. On the fovea ($\approx$ 1 degree), $\gamma_{max} \approx 0.5$ degree, the distance between centers of adjacent cones is about 22 arc sec, then the diameter of the fovea in number of cones is:

$$\text{fovea diameter} \approx \frac{1\,degree}{22\,arc\,sec}\, 3600 \frac{arc\,sec}{degree} \approx 164 \text{ cones}$$

In the present work, 80 photoreceptor elements will be used as the distance from the image plane center to the periphery of the workspace being analyzed. This distance is the *radius* of the central part of GRASP cameras (total CCD sensor device: 488 x 380 photoreceptors). As a simplification to plot the results of equations 2 through 7, a *square workspace* of side size equal to 160 pixels centered in the image plane is assumed. This range of x_p and y_p give the

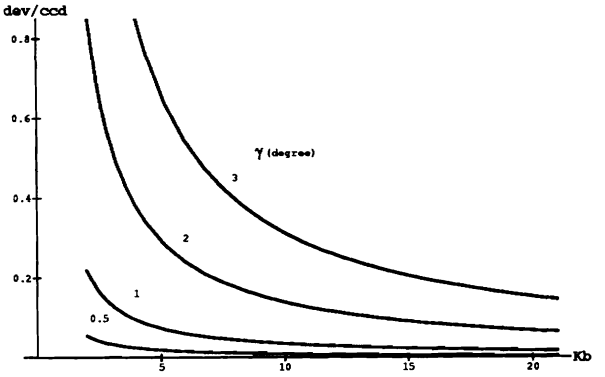


Figure 9: Projections deviation for GRASP platform system as a function of d_{obj} ($\varphi = 45^o$, b = 128 mm, f = 65 mm)

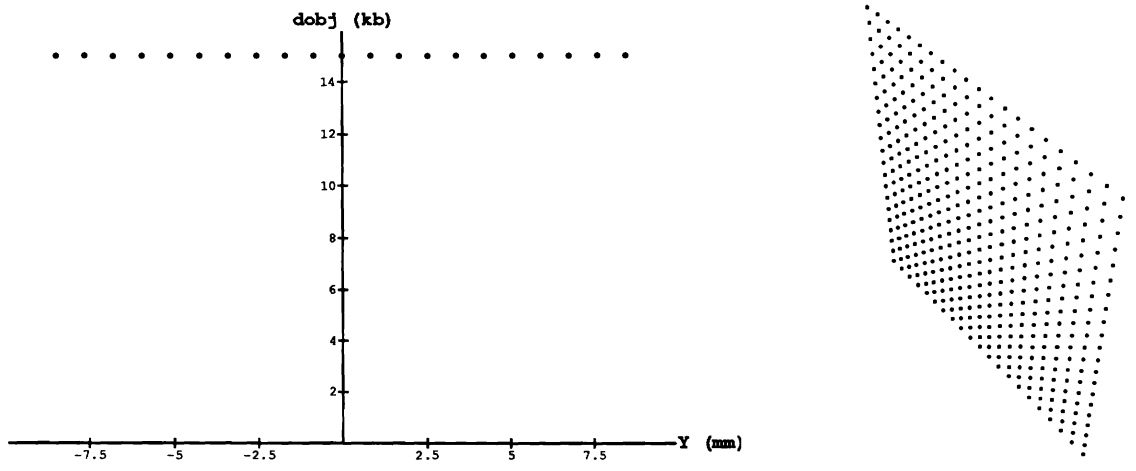
point horopter plotted in Figures 10 and 11, respectively for the human visual system and for the GRASP platform system. It can be seen from these Figures that the point horopter is a surface in the 3D space.

3 Micromovements

In the previous section the shape of the point horopter was analyzed for a given vergence angle calculated from the object distance under fixation. The main analysis now is conducted for a number of vergence angles α_i about the *fixation point in the visual plane.* The different values of α_i are composed by the following equation:

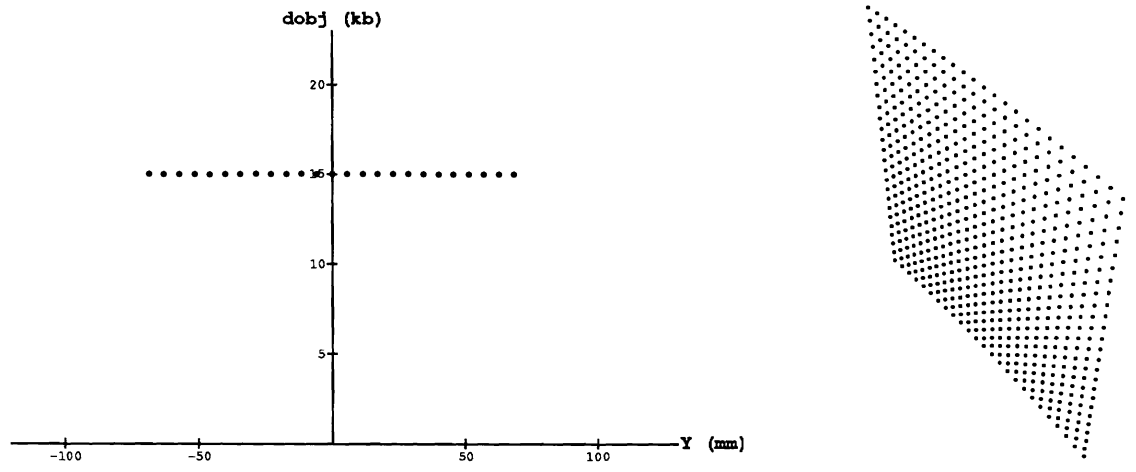
$$\alpha_i = \arctan(\frac{b}{2d_{obj}}) + \varepsilon_i \tag{18}$$

where, $\varepsilon_\mathbf{i}$ is a small angle increment (positive or negative) that describes the micromovement about the *fixation point* (first part of equation 18). The α_i set about a given fixation point determines a complete *micromovement cycle* about this point. The Figures 12 and 13 show the locus of "intersection" points in 3D space respectively for the human visual



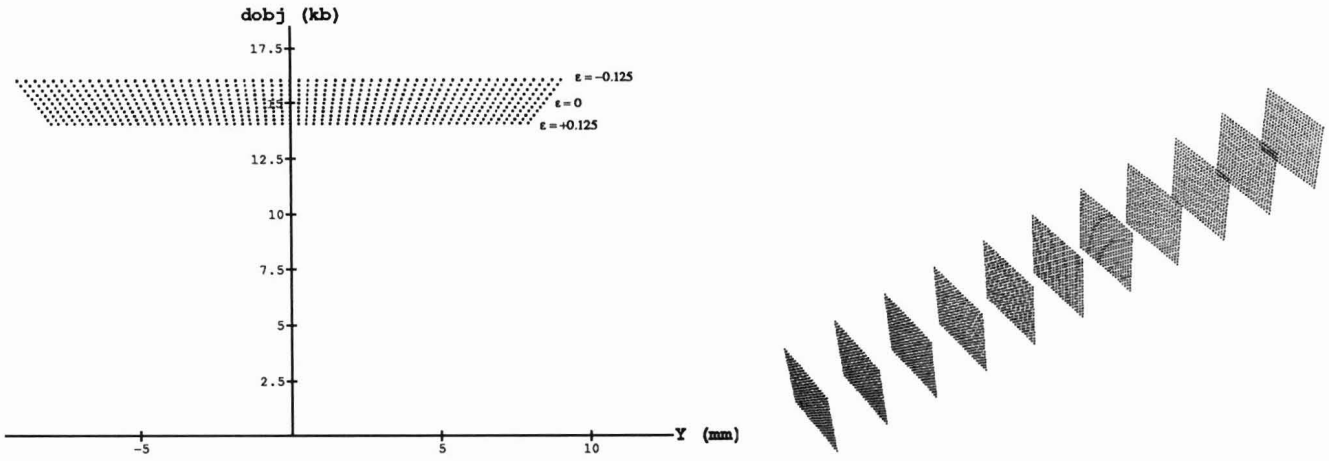
(a) (b)

Figure 10: Point horopter for human visual system (b = 65 mm, f = 17 mm, d_{obj} = 15 b, dce(0) = 1.81 10^{-3} mm, x_p and $y_p \in$ [-80 dce(0), +80 dce(0)]): (a) Top view, (b) Perspective view



(a) (b)

Figure 11: Point horopter for GRASP platform system (b = 128 mm, f = 65 mm, d_{obj} = 15 b, dce(0) = 30.0 10^{-3} mm, x_p and $y_p \in$ [-80 dce(0), +80 dce(0)]): (a) Top view, (b) Perspective view



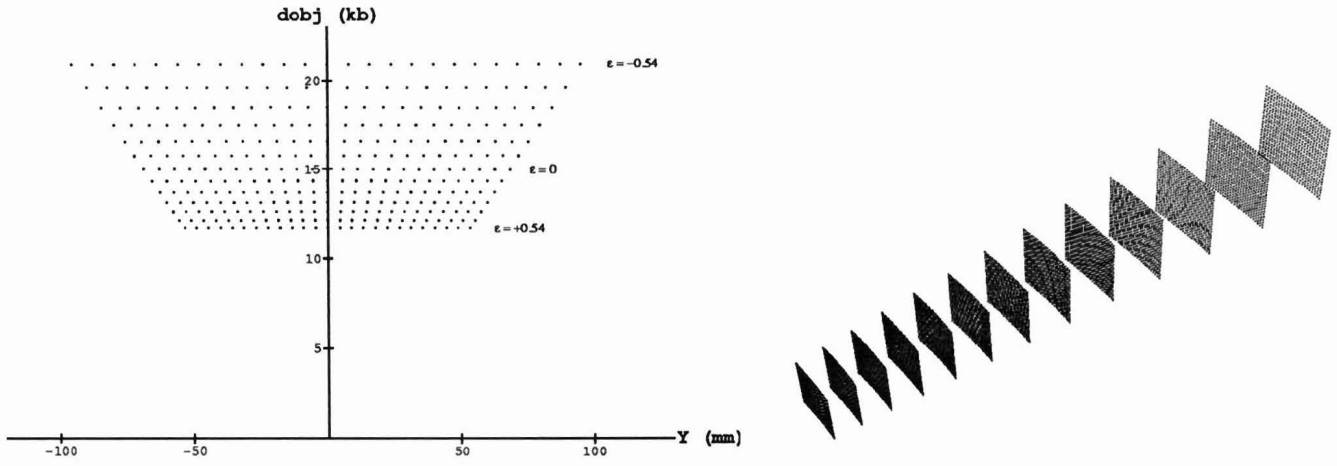
(a) (b)

Figure 12: Set of point horopter surfaces for human visual system (b = 65 mm, f = 17 mm, d_{obj} = 15 b, dce(0) = 1.81 10^{-3} mm, x_p and $y_p \in$ [-80 dce(0), +80 dce(0)], $\varepsilon \in [-0.125^o, 0.125^o]$): (a) Top view, (b) Expanded perspective view

system and GRASP platform system for a given micromovement cycle. The surfaces shown correspond to the set of point horopter surface generated for each vergence angle α_i.

As can be seen in Figures 12 and 13, the locus of all the "intersection" points form a volume in the 3D space. Therefore, 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.* Remember that this operator relates elements (image plane points) with the *same* horizontal and vertical distance from the center of the left and right image planes. It is possible to use a local correspondence operator since we presuppose that d_{obj} is greater or equal to fifteen baselines, implying a small *dev* (see previous section).

It is important to point out that errors in stereo, along z axis, with the present approach are due the vergence angle quantization (angle steps fixated by ε). These errors differ from the quantization errors due to discrete photoelements in cameras, that are a common characteristic of other stereoscopic methods. As described in [20], the errors due the photoreceptor quantization are significant and increase with the distance from the object to the cameras



(a) (b)

Figure 13: Set of point horopter surfaces for GRASP platform system (b = 128 mm, f = 65 mm, d_{obj} = 15 b, dce(0) = 30.0 10^{-3} mm, x_p and $y_p \in$ [-80 dce(0), +80 dce(0)], $\varepsilon \in [-0.54^o, 0.54^o]$): (a) Top view, (b) Expanded perspective view

system ([20] Figure 2: Diamond shaped regions). The present approach allows us to overcome the photoreceptor quantization limitation by using a sequence of pairs $[\alpha_i, \mathbf{C}_i]$, where α_i is the i^{th} vergence angle and $\mathbf{C}_i$ is the i^{th} matrix of correspondence responses. It must be emphasized that although the present analysis considers only the micromovements in the visual plane (horizontal micromovements), the real human eye system has micromovements in a vertical plane including the visual axis as well as rotations about the visual axis itself [6].

The application of the present approach in the visual system machinery permits the explanation of depth acuity (or stereoacuity). It is supposed that using *continuos vergence micromovements* the visual system can overcome the limitation of the photoreceptor (e.g. *cone* at fovea) dimension and it can detect depth with high accuracy. This topic will be treated in more details in the next sections.

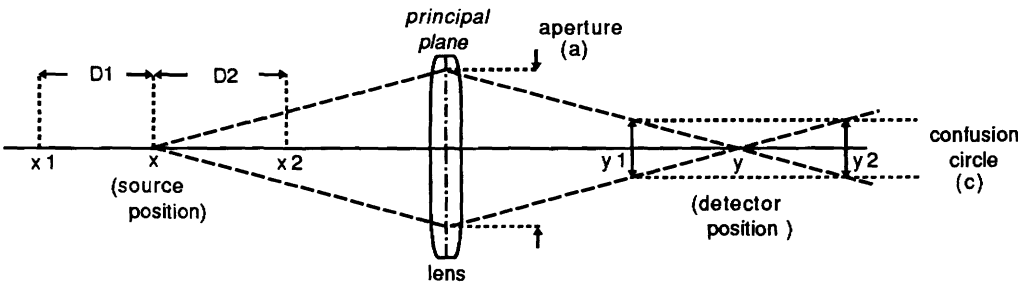


Figure 14: Depth of field

4 Depth of field

Because of the finite area of the photoreceptors elements on the CCD sensor, an object point may lie at different distances and still be imaged onto the same receptor [12]. For a circular lens, the defocused image of this object point is a circle called *blur circle*. Assuming that for any image the differences in sharpness cannot be distinguished for blur circles with diameter smaller than c, *circle of confusion*, the distance between the nearest and farthest object planes at which sharpness is obtained is called *depth of field* [5, 12]. Figure 14 (redrawn from [12], Figure 3.10) shows the notation for the following depth of field formula deduction:

$$\frac{c}{a} = \frac{y - y_1}{y_1} = \frac{y_2 - y}{y_2} \tag{19}$$

Applying the *Gaussian Lens* equation (1):

$$x_1 = \frac{x\ a\ f}{a\ f\ -\ c(x - f)} \tag{20}$$

$$x_2 = \frac{x\ a\ f}{a\ f\ +\ c(x - f)} \tag{21}$$

The *depth of field* can be deduced from (20) and (21) as:

$$D = D1 + D2 = \frac{2\ x\ a\ f\ c\ (x - f)}{a^2\ f^2 -\ c^2\ (x - f)^2} \tag{22}$$

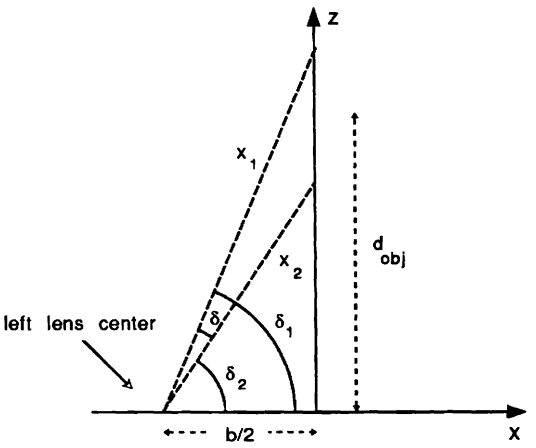


Figure 15: Micromovement cycle due to the depth of field in a stereo vision system

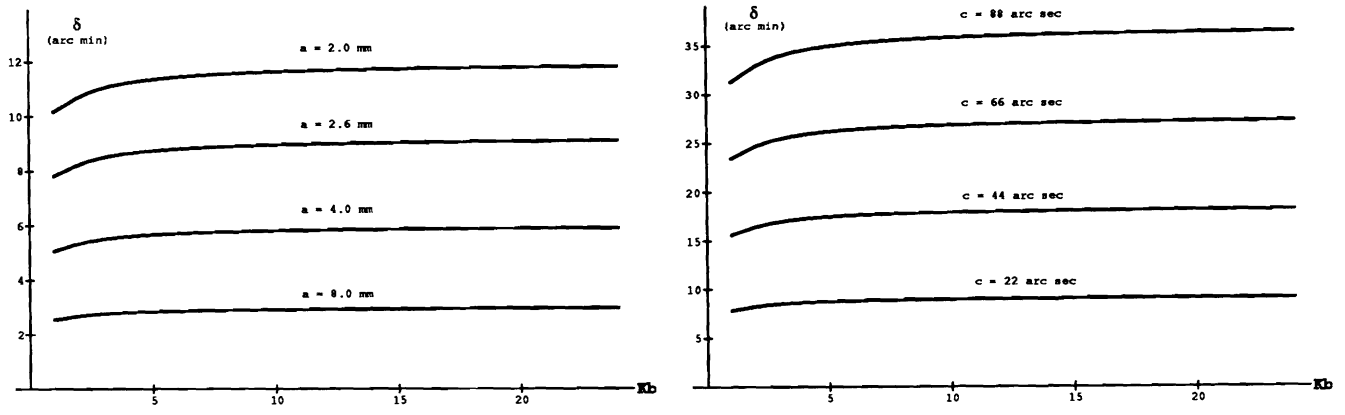
The present approach assumes that x_1 and x_2 limit the maximum range of the vergence micromovements about a given fixation point, i.e., they determine the maximum dimension of a complete *micromovement cycle* (section 3). Figure 15 shows the *micromovement cycle* (δ) as a function of x_1 and x_2. Using the notation shown in figure 15 we can deduce the upper limit of the micromovement cycle δ as being,

$$\begin{aligned} \delta &= \delta_1 - \delta 2 = \\ &= \arccos(\frac{b}{2\,x_1}) - \arccos(\frac{b}{2\,x_2}) \end{aligned} \tag{23}$$

where, x_1 and x_2 are given by the equations 20 and 21 with,

$$x = \sqrt{d_{obj}^2 + (\frac{b}{2})^2} \tag{24}$$

The application of equation 23 for the human visual system is shown in Figure 16. Note that the *upper limit of the micromovement cycle* due to the depth of field is approximately constant with the object distance and it varies with the aperture value and the confusion circle diameter.



(a) (b)

Figure 16: Micromovement cycle due to the depth of field for human visual system (f = 17 mm, b = 65 mm): (a) different apertures (c = 1.81 10^{-3} mm (= 22 arc sec)) (b) different confusion circle diameters (a = 2.6 mm)

5 The biological support of the micromovements

The following discussion of eye-movements in relation to physiological and psychophysical experiments is offered as a working hypothesis, useful for the understanding of the micro-movements role in depth perception.

Physiological results [11, 6] 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* from 1 to 20 arc min in amplitude, which have a velocity about 10° per second. The drift and flick movements, occur in opposing directions and produces convergence/divergence of the eyes [6]. The present approach supposes that *binocular flicks* are the basis for depth perception.

It is important to emphasize that the present approach refers to these micromovements as *binocular flicks* and not as *saccades* movements as opposed to [11], since [11] defines saccades

as *conjugate movements*. In conjugate movements the angle of intersection of the lines of sight of the two eyes remains constant. This is not the case of the *binocular flicks* reported in [6], since these occur in opposite directions in the two eyes producing vergence movements (disjunctive movements). The only similarity between *binocular flicks* and *saccades* is the high displacement velocity.

Some interesting results described in [6] are: the movements do not show significant difference when one eye is covered, and the flicks and drifts are very much larger when the fixation spot is extinguished. These results can be related with the dependence of the *binocular flicks* range with the depth of field.

Another experiment described in [6] appears to show that the eye is unable to see anything during the *flick* periods implying some vision inhibition mechanism. The explanation is that, during the *binocular flicks*, the visual system acquires only depth information of the image points prestored in memory.

Other group of experiments that support the present approach is obtained from electro-oculography. There is a small electrical potential between the cornea and the retina, which can be recorded by electrodes placed in pairs on the skin around the orbital rim [11]. With this electro-oculograms it is possible to analyze the optokinetic nystagmus (OKN) response elicited by moving stereoscopic contours. In [2] (figure 1) the determination of the presence or absence of OKN by the typical electro-oculographic test was unambiguous for a set of stereoblind (infantile strabismus) subjects. It concludes: "our results show a correlation of subjective responses and objective OKN responses in that subjects who are able to perceive random-dot stereogram target also give OKN response to moving stereoscopic contours while those who cannot perceive such target (stereoblind by definition) do not give an OKN response". Note that the subjects were stereoblind due to infantile strabismus implying that there is no possibility of local correspondence between elements on the same relative position at the left and right retinas. The present approach explains the absence of OKN by the lack of local correspondence.

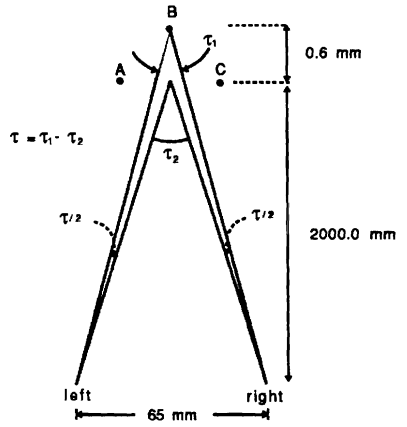


Figure 17: Angular disparity (τ) as in the Helmhotz experiment.

The above experiments give support to the approach that depth perception is acquired by the continuous micromovements, flicks, about a fixation point.

6 Integrating "flicks" with other biological evidences

Assuming the vergence micromovements mechanism as the basis of the depth perception, it is easy to understand the phenomena of *stereoacuity* (depth or stereoscopic acuity, stereopsis). As well described in [11] and [18], it is almost incredible that most observers under normal conditions can discriminate a difference in depth corresponding to an *angular disparity* (interocular disparity) of about 10 arc sec. The angular disparity (τ) is illustrated in Figure 17 which describes the scheme of the fundamental *three-needle* experiment devised by Helmholtz (see [11], pp. 484). 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% discrimination 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* ($\approx$ 22 arc sec).

The present approach explains this high sensitivity to slight disparity by the relation

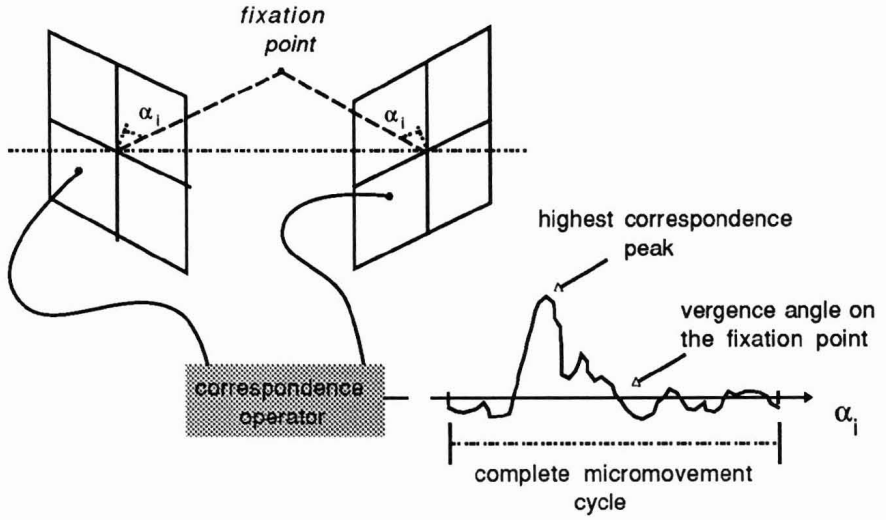


Figure 18: Correspondence operator response

between depth perception and the vergence eyes micromovements and not by the capability of the human visual system to spatially detect disparity on the retinas. The depth perception is obtained by the signal correspondence response of the left and right elements (e.g. cones at fovea) at the same relative position in the respective retinas (figure 18). Then 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 depth values by the highest peak of correspondence response during a complete *micromovement cycle* (section 3). 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 canceled for a given vergence angle.

Another phenomena that can be explained by the present approach is known in the literature ([11]) 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 un-

crossed (left and right optic axis do not cross) or crossed disparity (figure 19.a). The classical static limits for Panum's area, the mean crossed to uncrossed range of horizontal disparities, is reported by [16] as being 14 arc min. It must be emphasized that the experiments described in [9] 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 depth can be perceived from a disparity 500 times smaller.

According to [18] there have been four classic approaches to the binocular fusion: the synergy, local sign, eye movement, and the suppression hypotheses. There are enough justifications in [18] to discard the above hypotheses on the explanation of the binocular fusion together with the stereoscopic depth sensibility. The eye movement hypothesis of fusion described in [18] (Helmhotz hypothesis) suggests that small eye movements make the image so unstable that precise specification of stimulus position is impossible. This hypothesis is invalidated by the fact of the extremely fine resolution of stereoscopic depth (as described above).

In the present approach, the phenomena of binocular fusion and stereoscopic depth are assumed to be underlined 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* (section 3). 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 [6]. Note that the Panum's fusional horizontal radius must be compared to the total range of a *monocular* micromovement reported in [6] to be coherent to both definitions (figure 19). In the present analysis the *vertical* fusion radius is not considered since ([17]) this radius follows the monocular spatial resolution limit of the retina. 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.

The *classical* value of the Panum's area has being used in spite of the knowledge that

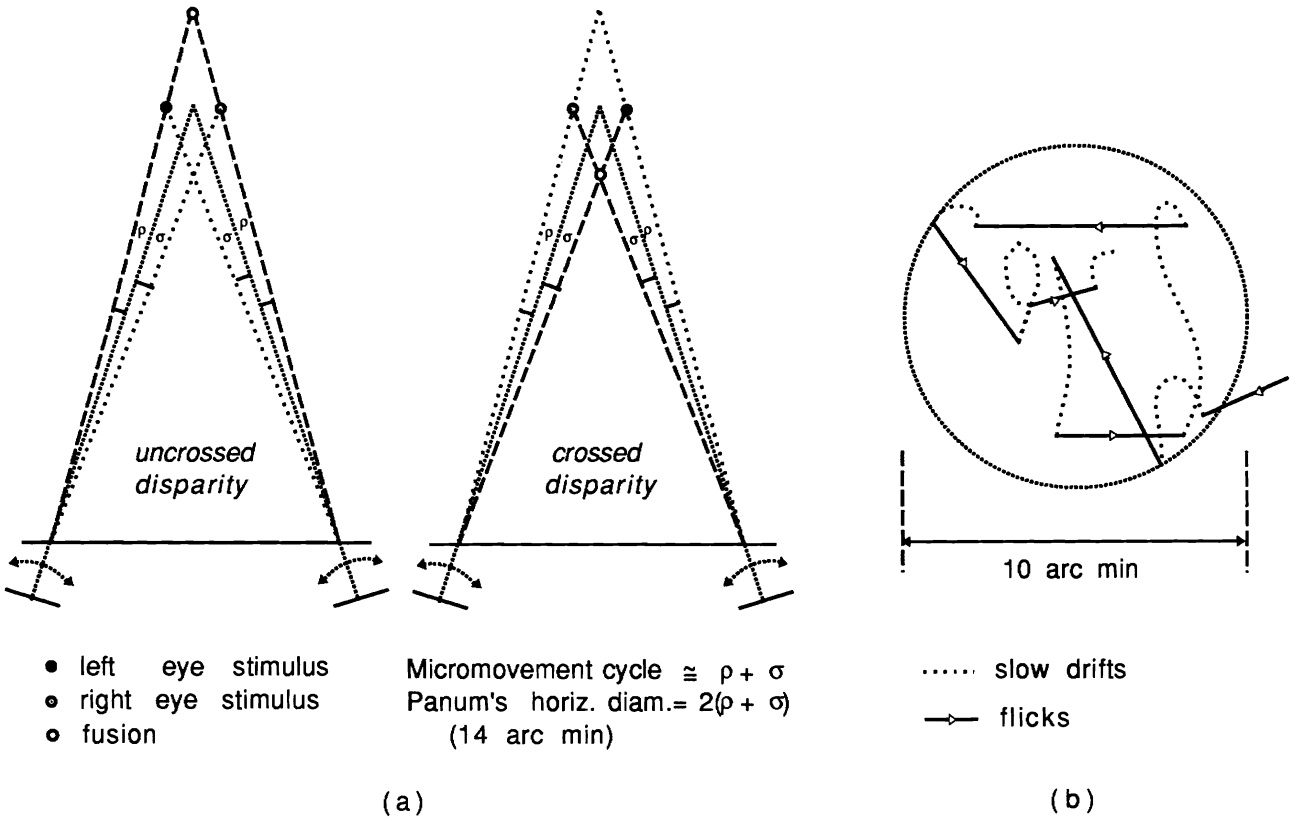


Figure 19: Micromovements cycle: (a) range due to the Panum's area diameter, (b) range due to the experimental measured micromovements (see [6] for details)

the limits of the Panum's area [19, 17, 7, 14, 9] can differ from the classical static zone of fixed extent. This restriction does not imply a loss of generality since the objective here is to introduce the relation between the Panum's area and the vergence eye micromovements.

Finally, with the supposition that the vergence micromovements mechanism underlies the binocular fusion phenomena, it is possible to explain why the Panum's area is greater for blurred targets than for focused targets as reported in [13]. As shown in section 4, Figure 15, the upper limit of the micromovement cycle depends on the values of aperture and blur circle. It is interesting to observe that for a blur circle diameter equal to the distance between adjacent cones in the fovea (c = 22 arc sec in Figure 15.b), the upper limit of the micromovements is approximately 7.0 arc min, that is, the same value of the classical Panum's fusional radius. Assuming that the human eye aperture (pupil size) is fixed, the upper limit of the micromovement cycle increases with the blur circle diameter (figure 15.b). It is known that the blur cycle is greater for blurred targets than for focused targets. Therefore, blurred targets imply a greater upper limit of the micromovement cycle and consequently a greater Panum's area can be expected.

7 Discussion

The present work does not specify the type of correspondence operator to be used in the generation of the correspondence response signal (see Figure 18). A simple difference operator has been used to realize the simulation. Although the difference operator is simple to implement it does not work well when an unequal monocular contrast exists between the left and right images. On the other hand this operator satisfies the differencing characteristic inherent in stereo, performed in the presence of highly correlated noise between the left and right images (see [14] for details). As future work, it is the intention to use a gradient detector operator of the grey-value function at every pixel in the left and right images. Each pair of left and right elements with the *same* horizontal and vertical distance from the center of the left and right image planes generates a correspondence response based on: the gradient direction, the

gradient magnitude and the weighted average of the grey-values about these elements. The above correspondence operator satisfies the observed experimental results [14] that the human visual system does not support unequal contrast between the two eyes.

As described in the previous section, the *continuous vergence micromovements* approach permits to overcome the physical limitation of the photoreceptor dimension (CCD element or cone) on the depth perception. There is another kind of quantization that limits the performance of the digital computer implementation of any vision system, which is the resolution (number of bits) of the analog/digital converter (A/D) that digitizes the video signals. This limitation can be overcome by implementing the *correspondence operator* in hardware (*correspondence unit*). In this way, the synchronized left and right video signals are the inputs of the *correspondence unit* and the output correspondence signal can now be digitized by the A/D. Therefore, during a complete micromovement cycle, the stereo system will acquire a sequence of pairs $[\alpha_i, \mathbf{C}_i]$, where α_i is the i^{th} vergence angle and $\mathbf{C}_i$ is the i^{th} "correspondence image" (correspondence matrix). A simple algorithm can now detect the highest correspondence peak and the respective α_i for each element of the correspondence matrix. Note that the *stereoscopic matching problem* still exists, since it is possible to have two or more correspondence peaks with similar values for an element of the correspondence matrix. It is not the intention here to suggest how to solve the matching problem. It is important to emphasize that, due to its local operation characteristics, the present approach can be implemented in a modular hardware scheme.

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 retina locus. Therefore, the concept of a "neural structure spread spatially" in the visual system to perceive depth via measurement of disparity is substituted by a "neural structure connected locally with the neighborhood" of each retina locus.

The *continuous vergence micromovements* approach is supported by experimental results from physiology and psychophysics. The experimental results from neurosciences (visual pathways, organization of the visual cortex, etc) and from psychophysics using flashed images as

well as the best correspondence operator, multiple direction detectors, multiple resolution, spatial frequency tuned channels, and so on will be the theme for future work.

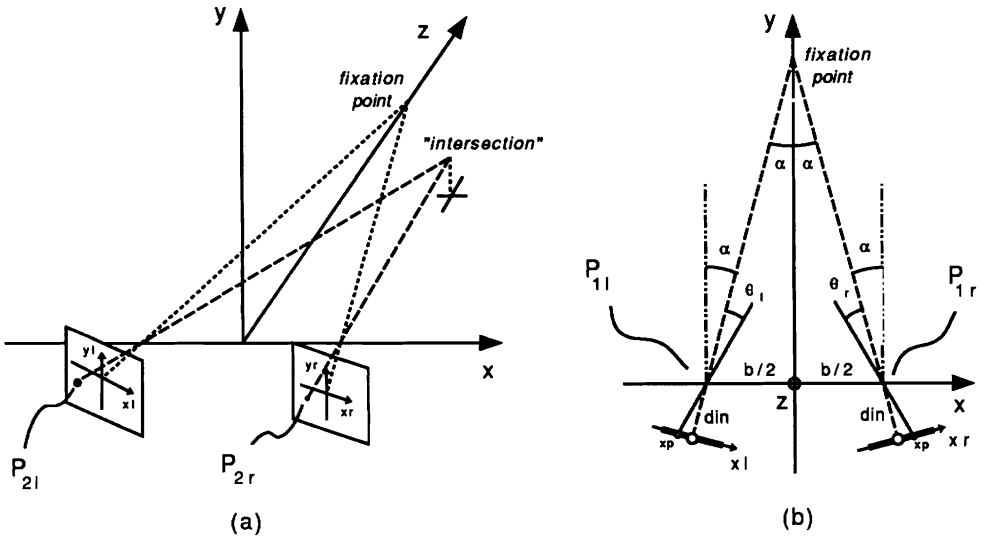


Figure 20: The vergence camera system: (a) Perspective view (b) Top view

A The vergence camera system equations

All the following expressions concern the symmetric fixation in the visual plane, i.e., the vergence angle of the two cameras have the same value α and the point beeing fixated is in the visual (horizontal) plane of the cameras. This fixation point is considered to be at the distance $\mathbf{d_{obj}}$ from the middle of the baseline. Associated with each image plane there is a coordinate system with its origin at the image plane center (figure 20). The lens has focal length **f** and their centers are separated by a baseline **b**.

Using the global coordinate system, according to Figure 20, the left optic axis is the line passing through the points $P_{1l}P_{2l}$ and the right optic axis is the line passing through the points $P_{1r}P_{2r}$, where

$$\vec{P_{1l}} = (x_{1l}, y_{1l}, z_{1l})^T = (-\frac{b}{2}, 0, 0)^T \quad (25)$$

$$\vec{P_{1r}} = (x_{1r}, y_{1r}, z_{1r})^T = (\frac{b}{2}, 0, 0)^T \quad (26)$$

$$\vec{P_{2l}} = (x_{2l}, y_{2l}, z_{2l})^T \quad (27)$$

$$\vec{P_{2r}} = (x_{2r}, y_{2r}, z_{2r})^T \quad (28)$$

$$(29)$$

Let α_l, for the left camera, be the vergence angle measured clockwise from the vertical line passing through P_{1l} and θ_l measured clockwise from $C_l P_{1l}$. Let α_r, for the right camera, be the vergence angle measured counter-clockwise from the vertical line passing through P_{1r} and θ_r measured counter-clockwise from $C_r P_{1r}$. Assuming that the analysis is restricted to symmetric fixation in the visual plane, with any position torsion and optical aberrations considered zero, the same vergence angle can be considered for both cameras, than $\alpha_l = \alpha_r = \alpha$. Also, with the analysis restricted to corresponding elements (same $(x_p, y_p)^T$) on both image planes, $y_{2l} = y_{2r} = y_p$, is valid for the following equalities:

$$x_{2l} = -\frac{b}{2} + x_p \frac{\sin(\alpha + \theta_l)}{\sin(\theta_l)} \quad (30)$$

$$z_{2l} = x_p \frac{\cos(\alpha + \theta_l)}{\sin(\theta_l)} \quad (31)$$

$$x_{2r} = \frac{b}{2} + x_p \frac{\sin(\alpha + \theta_r)}{\sin(\theta_r)} \quad (32)$$

$$z_{2r} = -x_p \frac{\cos(\alpha + \theta_r)}{\sin(\theta_r)} \quad (33)$$

where,

$$\alpha = \arctan(\frac{b}{2d_{out}}) \quad (34)$$

$$\theta_l = -\arctan(\frac{x_p}{d_{in}})$$

$$\theta_r = \arctan(\frac{x_p}{d_{in}}) = -\theta_l$$

To find the optical axes equations the two-point form for straight lines is used.

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

That implies,

$$y = (x - x_1)\frac{y_2 - y_1}{x_2 - x_1} + y_1 \quad (35)$$

$$z = (x - x_1)\frac{z_2 - z_1}{x_2 - x_1} + z_1 \quad (36)$$

Applying and reducing the above equations for the left and right cameras, the expressions that describe the locus of all 3D points on the left and right optic axis respectively are obtained. According to the notation shown in Figure 20, the following equations describe the left and right optic axes of a given corresponding element (same $(x_p, y_p)^T$) on the image planes coordinate systems.

For the left optic axis:

$$y = \frac{y_p}{x_p}(x + \frac{b}{2})\frac{\sin(\theta_l)}{\sin(\alpha + \theta_l)} \quad (37)$$

$$z = (x + \frac{b}{2})\frac{1}{\tan(\alpha + \theta_l)} \quad (38)$$

For the right optic axis:

$$y = \frac{y_p}{x_p}(x - \frac{b}{2})\frac{\sin(\theta_r)}{\sin(\alpha + \theta_r)} \quad (39)$$

$$z = (-x + \frac{b}{2})\frac{1}{\tan(\alpha + \theta_r)} \quad (40)$$

Note that for a given corresponding element (same $(x_p, y_p)^T$ on the image planes), it is valid that $\theta_r = -\theta_l$. Than the equations for the right optic axis become:

$$y = -\frac{y_p}{x_p}(x - \frac{b}{2})\frac{\sin(\theta_l)}{\sin(\alpha - \theta_l)} \quad (41)$$

$$z = (-x + \frac{b}{2})\frac{1}{\tan(\alpha - \theta_l)} \quad (42)$$

As can be observed in Figure 20, any optic axes passing through corresponding points on the image plane out off the *image coordinate system axes* do not intersect in the 3D space. This is due to the fact that the 3D "intersection locus" is necessarily closer to one camera than the other, resulting in a difference in the projection angles of the left and right optic axes. This difference in geometry of the optic axes obligate us to use some constraints to determine the coordinates of the "intersection" point $(x_\iota, y_\iota, z_\iota)^T$.

Since the present analysis is concerned with *zero horizontal* disparity, the 3D "intersection" point is determined for the same values of the x coordinate to the left and right optic axes. As we are interested in depth analysis, the other coordinate to be considered equal for both optic axes is the z coordinate. The above two conditions implies different values of y for the left and right optic axes at the "intersection" point. Therefore, the y coordinate of the "intersection" point is considered the mean of the left and right y coordinates.

Therefore, equating the z equations for both optic axes will result in:

$$x_\iota = \frac{b}{2}\frac{\tan(\alpha + \theta_l) - \tan(\alpha - \theta_l)}{\tan(\alpha + \theta_l) + \tan(\alpha - \theta_l)} \quad (43)$$

$$z_\iota = \frac{b}{\tan(\alpha + \theta_l) + \tan(\alpha - \theta_l)} \quad (44)$$

Substituting x_ι on the y equation:

$$y_\iota = \frac{1}{2}((\frac{y_p}{x_p}\frac{b}{2}(2\frac{\tan(\alpha + \theta_l)}{\tan(\alpha + \theta_l) + \tan(\alpha - \theta_l)})\frac{\sin(\theta_l)}{\sin(\alpha + \theta_l)}) + \quad (45)$$

$$(\frac{y_p}{x_p}\frac{b}{2}(2\frac{\tan(\alpha - \theta_l)}{\tan(\alpha + \theta_l) + \tan(\alpha - \theta_l)})\frac{\sin(\theta_l)}{\sin(\alpha - \theta_l)})) \quad (46)$$

Reducing the above expression, will result in:

$$y_\iota = \frac{y_p}{x_p}\frac{b}{2}\frac{\sin(\theta_l)}{\tan(\alpha+\theta_l)+\tan(\alpha-\theta_l)}\left(\frac{1}{\cos(\alpha+\theta_l)}+\frac{1}{\cos(\alpha-\theta_l)}\right) \quad (47)$$

$$= \frac{y_p}{x_p}\frac{b}{2}\frac{\sin(\theta_l)\ \cos(\theta_l)}{\sin(\alpha)} \quad (48)$$

But

$$\sin(\theta_l) = -x_p\frac{\cos(\theta_l)}{d_{in}}$$

Than

$$y_\iota = -\frac{y_p}{d_{in}}\frac{b}{2}\frac{cos^2(\theta_l)}{\sin(\alpha)} \quad (49)$$

□

Another set of equations are deduced to permit the analysis of the projections of a single point $\vec{P_o}$, at the position $(X_o, Y_o, Z_o)^T$ in the 3D space, on both image planes. The intersection of the left and right optic axes, passing through the left (P_{1l}) and right (P_{1r}) optic lens center respectively, with the image planes are determined. The intersection of the left optic axis with the left image plane is named $(x_{pl}, y_{pl})^T$. The intersection of the right optic axis with the respective image plane is named $(x_{pr}, y_{pr})^T$.

Using the equations 37 through 42, and substituting,

$$\sin(\theta_l) = -x_{pl}\frac{\cos(\theta_l)}{d_{in}} \quad (50)$$

$$\sin(\theta_r) = x_{pr}\frac{\cos(\theta_r)}{d_{in}} \quad (51)$$

It is easy to deduce the following equations:

$$x_{pl} = -d_{in}\tan(\arctan(\frac{\frac{b}{2}+X_o}{Z_o})-\alpha) \quad (52)$$

$$y_{pl} = \frac{-Y_o d_{in}\sin(\alpha+\theta_l)}{\cos(\theta_l)}(X_o+\frac{b}{2})^{-1} \quad (53)$$

$$x_{pr} = -d_{in}\tan(-\arctan(\frac{(\frac{b}{2}-X_o)}{Z_o})+\alpha) \quad (54)$$

$$y_{pr} = \frac{Y_o d_{in}\sin(\alpha+\theta_r)}{\cos(\theta_r)}(X_o-\frac{b}{2})^{-1} \quad (55)$$

where,

$$\theta_l = -\arctan(\frac{x_{pl}}{d_{in}}) \quad (56)$$

$$\theta_r = \arctan(\frac{x_{pr}}{d_{in}}) \quad (57)$$

□

A way to verify the above equations is to use a particular point in space $\vec{P}_o$ at the position $(x_\iota, y_\iota, z_\iota)^T$ generated by the "intersection" of the left and right optic axes passing through the same corresponding position $(x_p, y_p)^T$. The left and right projections of this point should imply *zero horizontal* disparity on image planes. This supposition can be confirmed doing:

$$\theta_l = -\arctan(\frac{x_p}{d_{in}}) \quad (58)$$

$$X_o = x_\iota = \frac{b}{2}\frac{\tan(\alpha+\theta_l)-\tan(\alpha-\theta_l)}{\tan(\alpha+\theta_l)+\tan(\alpha-\theta_l)} \quad (59)$$

$$Z_o = z_\iota = \frac{b}{\tan(\alpha+\theta_l)+\tan(\alpha-\theta_l)} \quad (60)$$

Applying equations 52 and 54, it can be verified that x_{pl} is equal to x_{pr} for the above values of X_o and Z_o. Therefore, there is no *horizontal disparity* between the projections on the left and right image planes.

It is important to emphasize that, for the above values of X_o and Z_o, y_{pl} is different from y_{pr}. This was expected since the "intersection" locus in space is necessarily closer to one camera than the other, resulting in a difference between the projection angle in the two cameras.

References

[1] A. L. Abbot and N. Ahuja. Surface reconstruction by dynamic integration of focus, camera vergence, and stereo. In *Proc. Second Intl. Conf. on Computer Vision*, pages 532–543, Tarpon Springs, FL, December 1988.

[2] S. M. Archer, K. K. Miller, and E. M. Helveston. Stereoscopic contours and optokinetic nystagmus in normal and stereoblind subjects. *Vision Res.*, 27(5):841–844, 1987.

[3] R. Bajcsy. Active perception vs. passive perception. In *Proc. Workshop on Computer Vision*, pages 55–59, Bellaire,MI, October 1985.

[4] R. Bajcsy. Perception with feedback. In *Proc. DARPA Image Understanding Workshop*, pages 279–288, Cambridge, MA, April 1988.

[5] S. Das and N. Ahuja. Integrating multiresolution image acquisition and coarse-to-fine surface reconstrution from stereo. In *Proc. IEEE wksh on Interpretation of 3D Scenes*, 1989.

[6] R. W. Ditchburn. Eye-movements in relation to retinal action. *Optica Acta*, 1(4):171–176, 1955.

[7] Fender D. H. and Julez B. Extension of panum's fusional area in binocularly stabilized vision. *J. opt. Soc. Am.*, 57(6):819–830, 1967.

[8] R. N. Haber and M. Hershenson. *The psychology of visual perception.* Holt, Rinehart and Winston, Inc., 1973.

[9] T. Heckmann and C. M. Schor. Panum's fusional area estimated with a criterion-free technique. *Perception & Psychophysics*, 45(4):297–306, 1989.

[10] W. Hoff and N. Ahuja. Surfaces from stereo: integrating feature matching, disparity estimation, and contour detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(2):121–136, 1989.

[11] J. W. Kling and L. A. Riggs. *Experimental psychology.* Holt, Rinehart and Winston, Inc., 1971.

[12] E. P. Krotkov. *Exploratory visual sensing for determining spatial layout with an agile stereo camera system.* PhD thesis, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, 1987.

[13] J. J. kulikowski. Limit of single vision in stereopsis depends on contour sharpness. *Nature*, 276:126–127, 1978.

[14] G. E. Legge and Y. Gu. Stereopsis and contrast. *Vision Res.*, 29(8):989–1004, 1989.

[15] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. R. Soc. Lond.*, B(204):301–328, 1979.

[16] D. E. Mitchell. A review of the concept of panum's fusional areas. *Am. J. Optom.*, 43:387–401, 1966.

[17] C. Schor, I. Wood, and J. Ogawa. Binocular sensory fusion is limited by spatial resolution. *Vision Res.*, 24(7):661–665, 1984.

[18] C. M. Schor and K. J. Giuffreda. *Vergence eye movements: basic and clinical aspects.* Butterworth, 1983.

[19] C. M. Schor and C. W. Tyler. Spatio-temporal properties of panum's fusional area. *Vision Res.*, 21:683–692, 1981.

[20] F. Solina. *Errors in stereo due to quantization.* Technical Report MS-CIS-85-34, GRASP LAB Dept. of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, 1985.

[21] C. V. Stewart. New results in automatic focusing and a new method for combining focus and stereo. In *Proc. SPIE*, 1989.