Dynamic Fixation

K. Pahlavan, T. Uhlin and J-O. Eklundh*

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

Abstract: *Fixation is the link between the observer and the events in the outside world. It is also the physical action due to the attentional mechanisms embedded in an active vision system. Therefore, fixation is a key element in the control strategy of an active vision system. Although fixation is not necessarily a binocular process, it has been considered as such in computer vision. We will discuss the components of the human fixation model, and through its geometry and organization we will describe a machine fixation model, running in real-time on our head-eye system.*

1 Introduction

The presence of anthropomorphic head-eye systems has for the first time made it feasible to investigate the practical and computational benefits of active vision in general and primate vision in particular. Now, there exists a possibility to experiment on the practical value of the approaches based on biological constructions.

In human vision, the process of point to point binocular fixation has two components—vergence and version. The vergence component comprises accommodative vergence and disparity vergence. The relation between these two components of vergence is very interesting. At long distances where the change in blur is not significant, the pattern of the object seen by the two eyes are very similar and therefore the disparity detection is much easier. At short distances, the relation is reversed; i.e. the blur changes are significant while the matching procedure becomes less reliable. In this article, we will not discuss the mechanism behind version. We will rather emphasize on vergence in isolation, i.e. we assume that our system has already completed its binocular saccade towards the direction of interest.

We will briefly describe how the fixation process functions in human beings in general, then one of the essential components of fixation—vergence—is studied in isolation. Finally some experiments on the KTH-head are presented. The objective of the paper is, motivated by the needs for gaze control in head-eye systems, to clarify some fundamental relations in primate vergence movements—the relations between disparity and blur on the one hand and vergence and accommodation on the other hand. Following the geometry behind vergence movements, we also suggest that the mechanism for disparity detection can use the symmetry present in the pattern of the left and right images through isolating vergence from version.

*This work is supported by The Swedish Research Council for Engineering Science (TFR) and The Swedish National Board for Industrial and Technical Development; this support is gratefully acknowledged. We also thank Akihiro Horii for his implementation and study of focusing and blur detection

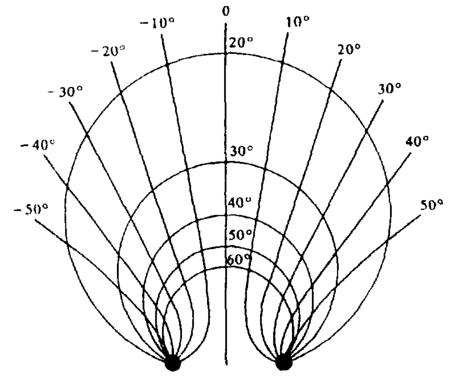

Figure 1: The lines of equal vergence (the circles through the eyes) and the lines of equal mean version (the hyperbolae through the eyes). From [3] after [13].

2 Fixation in primates

In human beings, vergence and version are integrated in a complicated manner so that one could not simply claim their relative independence of each other. A strict division into pure version and pure vergence, however, elucidates the geometry behind the mechanisms of the two movements. Observations from studying eye movements in human subjects have led to a model for such a geometry. A model of this kind is illustrated in Figure 1. In this figure, the lines of equal version form a series of rectangular hyperbolae that pass through the center of rotation of the two eyes and whose center is the midpoint of the baseline. The lines of equal vergence are represented by the Vieth-Müller circles passing through the center of rotation of the two eyes.

The two movements associate with two kinds of disparities. Pure version associates with zero disparity and pure vergence with symmetric disparity; these two qualities are especially interesting for our work. The version component is the ballistic saccadic movement due to zero disparity. Contrary to version, the vergence movement is image driven and relatively slow, due to two different stimuli; blur and symmetric disparity.

0-8186-3870-2/93 $3.00 © 1993 IEEE

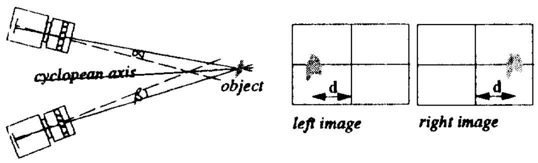

Figure 2: The geometry of vergence. The objects located along the cyclopean line are symmetrically located from the fixation point in time (the center of the images), i.e. $\alpha = \beta$.

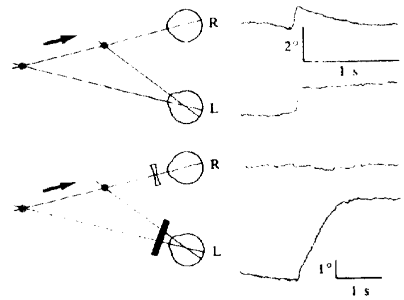

Figure 3: Under normal binocular viewing, the vergence/accommodation movement is symmetrical even if the object is moved along the line of sight on one of the eyes (above). Covering one eye (removing disparity) and introducing blur to the eye that has the object along its line of sight, results in a monotonic asymmetric movement of the eye. From [3] after [1, 2].

2.1 The geometry of vergence

Study of vergence in isolation from version leaves us a simple geometry. This geometry is illustrated in Figure 2. One could imagine a fictitious line of sight representing the common stimuli to the two eyes. This line is usually called the cyclopean axis and is one of the hyperbolae depicted earlier in Figure 1. The combination of the stimuli from the two eyes implies that depending on the contribution from each eye, the eyes could have complementary dominance in relation to each other. In the case when the two eyes are equally dominant, the image of the same object in the scene will be symmetrically located with the vertical axis through the fixation point as the symmetry axis.

The mentioned cyclopean axis plays a central role in the process of fixation, and vergence is defined by the displacement of the fixation point along this very axis. The angles α and β in Figure 2 are equal[1]; this is due to the fact that the cyclopean axis is actually a hyperbola and not a straight line. The equality of the two angles α and β is a very interesting quality. All the objects lying on the cyclopean axis are symmetrically projected into the cyclopean image. That is, *the image of the objects that are located along the common line of sight, are found on different sides of the left and right images and the eccentricity of their position in the images is a function of the distance of the projected object from the imaging system.*

With the cyclopean image, we mean the common representation for the left and right images which contains the sense of depth in some way. In the human case, it is known that the fusion of the left and the right images is not total. The two images are fused only in Panum's fusional area. This area, however is small compared to the field of view common to both eyes. Although we normally do not notice this, we are actually seeing most of the objects in our environment as doublets. Despite this fact, we have a very good sense of direction that should also be considered as a part of the cyclopean representation. Besides, there are usually plenty of other attentional stimuli, e.g. sound, changes in the scene, etc. that are sensed as directions rather than positions. Once the direction of interest is known, the problem will be how to find the source of interest, i.e. its position; this is when vergence completes the process of fixation.

[1] Of course, this is true only if the two eyes are equally dominant. Otherwise, the angles follow the same proportionality.

3 Vergence and its stimuli

In this section we consider two sources of stimuli for vergence. The vergence process in primates is known to be stimulated by both blur and disparity. Therefore, it is perhaps easier to see vergence as two different cooperating and competing processes using different feedback as stimuli; these two processes are accommodative vergence and disparity vergence.

3.1 Accommodative vergence and blur

In 1826, J. Müller showed in a classical experiment, that vergence movement in human vision is reproducible even in the absence of disparity [17]. The procedure is depicted in Figure 3. It is obvious from the experiment that changes in blur (even though blur is a monocular stimulus) result in a vergence movement with the perceiving eye as the totally dominant one. Other experiments point at the linear relationship between the angle of convergence and the accommodation stimulus [3].

Blur can be determined by an analysis of the spatial frequency contents of an image. Currently, the widely utilized CCD sensors have a rather low resolution (compared to human fovea) and therefore, they are not capable of detecting small changes in the frequency domain. That is, given the present facilities, focusing and blur detection is a well-posed problem up to a certain limit. Another way of describing this limitation is to state that the accommodation process succeeds in its task depending on the local density of texture vs. sensor resolution.

Another limitation imposed on accommodation is depth of field which is a function of both sensor resolution and iris opening; the smaller the depth of field is, the more precise the depth estimate becomes. In the human case, the very fine resolution of retinal cones in fovea centralis guarantees that at the final steps of

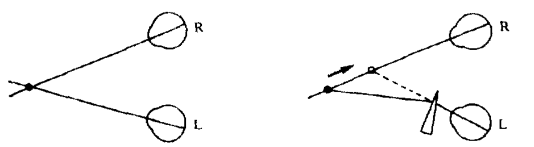

Figure 4: Introducing a weak prism in front of one eye when the eyes are fixating a point results in a unilateral vergence to bring the two retinal images together. From [3].

accommodation adjustments, a rather precise depth estimation can be achieved.

Static approaches to the problem of depth from blur usually use some constraints based on assumptions about what the best focus is (see e.g. [23] where a step edge is assumed to be a vertically abrupt discontinuity). This model, however, does not agree with real edges and discontinuities; resulting in a wrong depth estimate in real images. A good way to go around the problem is to use continuous feedback and keep track of how the sharpness enhancement develops. That is, the focusing should be an iterative procedure rather than a purely predictive one. This is applicable both on the process of depth from focus and depth from defocus. Work in our laboratory [6, 7] and experiments by other researchers [12] confirm the procedure above.

3.2 Disparity vergence

In Section 3.1, we mentioned Müller's experiment that demonstrated how blur, despite being a monocular cue resulted in a vergence movement. However, disparity is a binocular cue and is defined as the angle of correspondence of two associating patterns in the left and the right eye respectively. The mentioned experiment was evidence for sufficiency of blur as a vergence stimulus. Similar experiments by [4, 5, 11, 15, 16][2] show that the disparity stimulus is a sufficient cue for vergence as well. They also show that it is possible to simulate situations where pure disparity information can suppress the accommodative stimulus and even affect the crystalline lens and the pupil size. The procedure is depicted in Figure 4 where an insertion of a weak prism in front of one eye, without affecting the accommodation distance, results in a change of fixation.

It was described earlier that lack of knowledge about how sharp a well focused pattern is, results in a sequential procedure for accommodation; the sharpest pattern is chosen in this strategy. This is also the case with disparity detection. This issue deserves more elaboration.

Disparity detection is preconditioned by successful matchings. The problem is that, to our knowledge, there is no ***absolute*** measure for a good match; especially without presence of distinct features. This is in turn due to the fact that two corresponding patterns in a stereo pair are generally only ***qualitatively*** similar. A short baseline and a long distance to the object yield a better correspondence.

[2]The established experiments in physiology are cross references from [3] and [22].

In an active vision system, the uncertainty in disparity detection can be decreased by keeping track of the symmetrically detected matches under the process of convergence or divergence. Although we are only interested in symmetrically located similar patterns in the cyclopean image, there could exist objects in the scene that generate almost similar patterns in the two images as if they were projections from one object positioned along the cyclopean axis.

In order to take advantage of the resident information, blur should be computed. Earlier, we talked about the difference between the active approach and the passive approach in computing blur. An active vision system considers blur as a relative measure, i.e. as a measure for the sharpness degradation in relation to the accommodated point at the time.

The measure of blur is the high frequency contents of the area in the image where the blur analysis is to be done. This calls for some kind of Fourier analysis or equivalent shortcuts. Since blur is a relative measure, it is necessary to have a reference value for comparing the relative changes. Consequently, the discussion begins with sharpness criteria and continues with finding the reference point—the point of best sharpness. In the end of the section a computational model for detecting blur and depth from defocus will be suggested.

3.3 Sharpness criteria

What blur means is that the illumination from a point is spread over a blur circle rather than projected on a sharp point. We have examined several sharpness criteria in our laboratory; following earlier work by [21, 9, 12] and many other researchers. The performance of the focusing algorithms based on these criteria can be found in [6, 7]. The sharpness criterion we have used in the experiments described in this paper is the variance criterion. Since the sharpness of the discontinuities raised by the features in an image result in a large grey-level variance, it is rational to use the grey-level variance as a sharpness criterion. The formulation of variance in this case could look like:

$$\sigma^2 = \frac{1}{N^2} \sum_{x=1}^{N} \sum_{y=1}^{N} (I(x,y) - \mu)^2 \qquad (1)$$

where μ is the mean of the gray-level distribution. The maximum acquired value of σ^2 is then the criterion for good sharpness.

3.4 Blur detection

It is quite often desirable to have a measure of blur rather than achieving sharpness. The difference between a sharpness criterion and a blur measure is that the sharpness criterion is based on finding the best

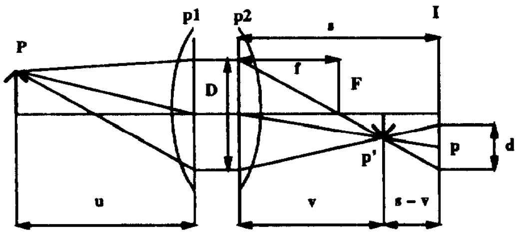

Figure 5: **The camera geometry and camera parameters; p1: first principal plane, p2: second principal plane, P: object point, p: image point, I: image detector plane, s, f, D: camera parameters (s: distance between the image detector plane and the second principal plane, f: focal length of the lens, D: numerical aperture of the lens), u: the object distance, v: screen plane distance, d: blur circle diameter**

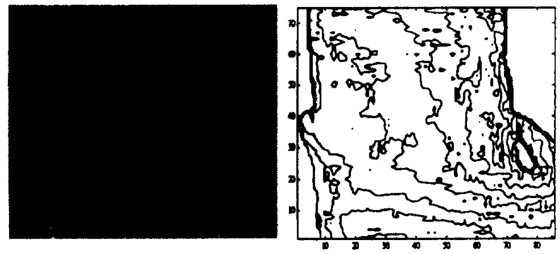

Figure 6: **The image of a chair and the level curve representation of the result obtained by active approach to depth from defocus. The chair is at a distance of about 2 meters. Two frames are used here, and the difference between the two images is a small change in the accommodation distance. Each level curve stands for a relative distance of 100 mm. For detailed information see [7].**

sharpness and therefore applicable when the sharpness is achieved sequentially; the blur measure, on the contrary, can be used to estimate the depth without completed accommodation.

The blur of an image due to defocusing can be described by a point spread function. Let P be a point in the scene and p be its focused image (see Figure 5). If P is not in focus then the image of P becomes a circular image called the *blur circle*. The point spread function is a function which represents the precise structure of the blur circle. The blurring of the image due to defocusing is considered to be described by the point spread function.

The relation between the position of P and p is given by the lens formula. From the geometry of the optics depicted in Figure 5 and the lens formula, the diameter of the blur circle d is given by:

$$d = Ds\left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right) \tag{2}$$

The blur circle can be described by the following two-dimensional Gaussian function ([19, 20]):

$$h = \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}}; \sigma = kd = kDs\left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right) \tag{3}$$

k is a calibration factor. Let $g(x, y)$ be the observed image of an object on the screen, and $f(x, y)$ be the corresponding focused image. Also, let $G(\omega, v)$ and $F(\omega, v)$ be the corresponding Fourier transforms. Then,

$$g(x, y) = h(x, y) \otimes f(x, y); \; P(\omega, v) = e^{-(\omega^2+v^2)\sigma^2}FF^* \tag{4}$$

The degree of defocusing is estimated by the difference of the focused image and the blurred image in the frequency domain.

A scenario for the active approach in estimation of depth from blur is manipulating the blur parameter σ for an image region by varying one or more of the camera parameters: s, f, and D (see equation (3)). Here, we recover depth by varying s, the focusing position of the lens, by small values.

Let two successive images be acquired with camera settings s and $s + ds$ where ds is a small change in the respective parameters. From equation (3) we have:

$$d\sigma = kD\left(\frac{1}{f} - \frac{1}{u}\right)ds \tag{5}$$

From equation (4):

$$dP = -2(\omega^2 + v^2)P\sigma d\sigma; \sigma d\sigma = -\frac{1}{2}\frac{1}{\omega^2 + v^2}\frac{dP}{P} \tag{6}$$

From (3), (5) and (6) we obtain

$$k^2D^2\left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right)\left(\frac{1}{f} - \frac{1}{u}\right)ds = -\frac{1}{2}\frac{1}{\omega^2 + v^2}\frac{dP}{P} = C$$

replacing $K = k^2D^2ds$ and $X = \frac{1}{f} - \frac{1}{u}$ and solving the equation for u with the constraint $f < u < \infty$:

$$u = \frac{1}{\frac{1}{f} - \frac{K+\sqrt{K^2+4KCs^2}}{2Ks}} \tag{7}$$

This way the resident information in the blur (the distance u) is extracted. It should, however, be underlined that the depth information obtained from blur and sharpness are very qualitative; this is especially true in the case of blur. The approach above is applied to a sequence of two images, one of which is depicted in Figure 6. The other image is identical to this one except for a slight change in the accommodation distance. The relative change in the blur is detected in the two images and the level curve depth map in the figure is generated.

4 Computing disparity

Correspondence is usually detected by finding the affinity of the patterns in the two images—matching. Matching can be performed in different ways and on different kinds of features. Nevertheless the fundamental problem is to find the similarity of the patterns. In the particular application that is the experimental basis for this paper, we have used normalized correlation method for finding the best match in the two images. Our objective is not discussing good disparity detection algorithms, but discussing the relation between accommodative and disparity vergence.

The normalized correlation is performed so that for each region on the left side of an image, there is another region on the right side of the other image that could contain the right pattern. If the correlation between these two regions yields a high peak, the match is accepted. The normalized correlation is formulated as:

$$C_{xy} = \frac{Cov[L, R]}{\sqrt{Var[L]Var[R]}}$$

$Cov[L, R]$ is the covariance and $Var[L]$ is the variance.

5 Integrating blur and disparity

In the end of last century there was a belief that accommodative vergence was the controlling component of vergence and the disparity vergence had only a supplementary role in the process [14]. Later experiments (like the one mentioned in 3.2 and illustrated in Figure 4) show, however, that the relation is more complex.

Figure 7 illustrates how the two blur and disparity stimuli can be integrated to realize a mutual effect from both stimuli on both accommodation and vergence. As long as the two stimuli agree, i.e. the two stimuli increase or decrease together, small errors do not matter, because the process is a closed loop one and as such, it can overcome small amounts of noise. However, the two stimuli result in two sets of actions that can be seen as two different processes—accommodation and vergence. There are situations where these two processes come into a conflict situation. Note that we are now talking about the situations where the *processes* are in conflict and not the stimuli. We have already mentioned the classical experiments where the researchers have isolated one stimuli and analyzed its effect on accommodation and vergence.

The relationship between accommodation and vergence can be distinguished into four categories:

- The two processes agree and both of them are wrong.
- The two processes agree and both of them are right.
- The two processes disagree since accommodation is confused.
- The two processes disagree because disparity is not unique.

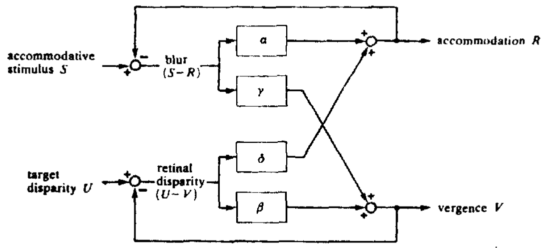

Figure 7: A model of the relationship between accommodative stimulus (S), target disparity (U), accommodation (R) and vergence (V): $R = \alpha(S-R) + \delta(U-V); V = \beta(U-V) + \gamma(S-R)$. From [3].

One could wonder how two processes that use the same stimuli can disagree. The answer is that they would not, if the disparity and blur stimuli were unique. Later in this section, we will discuss two cases where the stimuli are misleading the processes.

If we only have these two processes to integrate[3], then the first item would lead to a system reset, and nothing can be done about it; in practice the total darkness is an example of this case when the eyes move to the resting situation (focusing at infinity). In the second case every thing is just fine, since the system is a closed loop one and manages to filter out the small but inevitable noise.

The third and the fourth cases are the interesting ones and are discussed in the following subsections.

5.1 Erroneous detection of disparity

An erroneous vergence could be due to two reasons—repeated patterns and improper matches. An example of the first case is looking at wall papers or a fence from a near distance so that most of the retina is covered by the repeated pattern. Such an example is addressed in [18]. In this case the accommodation process overrides the vergence process.

An example of the second case is when no corresponding matches are found. This is the case when one of the cameras is occluded. Implicitly, this means that there already exists an interesting pattern or feature that is chosen in the image of the eye which is not occluded. That is the seeing eye becomes totally dominant and the disparity stimulus does not contribute to the accommodation process. The blur stimulus becomes dominant in the cue integration and the accommodation process overrides vergence.

5.2 Confusion in accommodation

The confusion in the estimation of accommodation distance can be categorized in two cases. Since accommodation is basically a monocular process, it is perhaps not correct to talk about accommodation *error*.

[3]This is not true in the case of primates. We have already mentioned that even pupil size is involved as a parameter in this context. The major task of the iris, however, seems to be regulation of the amount of light falling into the eyes.

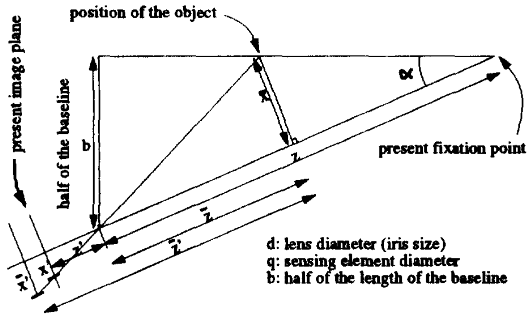

Figure 8: **The geometry behind the computation of depth of field and its relation to the area in the image where the disparity is the only cue.**

It is more proper to talk about the confusion raised by the internal conflict in the binocular accommodation, i.e. the conflict between the right and the left eyes' accommodations. An example of this is occlusion of one eye.

As the case was with the disparity error due to occlusion, the eye which is not occluded should become totally dominant, unless the occluding object, or other objects free from the occlusion are chosen to be fixated at. In the latter case, a version is involved which is not covered by our discussion here.

The other case is when the common accommodation is correct, but the binocular accommodation distance does not agree with vergence. An example of this case is looking at stereograms which puts accommodation and vergence in conflict; still fusing stereogram images does not affect the sharpness of the images, i.e. the accommodation process per se is intact and functional; it is the perceived depth which is the conflict issue. Besides, the accommodation process receives erroneous disparity stimulus. Nevertheless, it is obvious that most people can fuse the images. An important factor here is that the disparity cue is unique and very stable in all stereograms[4]. Furthermore, it should be noted that without special glasses, the fusion is unstable and only can be preserved voluntarily.

6 Computing depth of field

The dependency of disparity on blur can be expressed by saying that the disparity detection can be limited to the area in the image associated to the depth of field of the imaging system. In this sense it is very important to know where this area in the image is and to what extent should the matching be done. In the small area defined by the depth of field, it is only the disparity alone which takes the whole responsibility for finding the correspondences.

[4] See for example the many experiments described in [10].

Figure 8 illustrates the geometry of a lens needed for the rest of the discussion in this section. From the figure we have $\tan\alpha = \frac{\bar{x}}{z-\bar{z}}$, $\sin\alpha = \frac{b}{z}$. If α is small (which means that the paraxial approximation also valid):

$$\bar{x} = b\frac{z-\bar{z}}{z} \quad and \quad \frac{\bar{x}}{\bar{z}} = \frac{\bar{x}'}{z'} \to \bar{x}' = bz'(\frac{1}{\bar{z}} - \frac{1}{z}) \qquad (8)$$

According to [8], the depth of field can be defined as: $\frac{d}{z'}|\bar{z}' - z'| < q$. This along with the lens formula yields:

$$d|\frac{1}{f} - \frac{1}{z} - (\frac{1}{f} - \frac{1}{\bar{z}})| < q(\frac{1}{f} - \frac{1}{\bar{z}}); d|\frac{1}{\bar{z}} - \frac{1}{z}| < q(\frac{1}{f} - \frac{1}{\bar{z}})$$

For $z > \bar{z}$ we will have:

$$(d+q)\frac{1}{\bar{z}} < q\frac{1}{f} + d\frac{1}{z}; (d+q)(\frac{1}{\bar{z}} - \frac{1}{z}) < q(\frac{1}{f} - \frac{1}{z}) = q\frac{1}{z'}$$

In equation (8): $\bar{x}' < b\frac{q}{d+q}$ and for $z < \bar{z}$ equivalently. Finally considering $d >> q$

$$|\bar{x}'| < \frac{bq}{d}$$

that is, the area in which disparity is the only cue to a correct match, is restricted to bq/d and therefore, it is only a function of the size of the baseline, the iris and the size of the sensor elements.

7 Experiments

To illustrate the principles described earlier, we demonstrate the results from three sets of experiments here. In the first experiment, the disparity patterns are clear and the object is chosen so that the blur contribution is not directly decisive. In the second experiment, the patterns are much more complicated and natural; there is a major risk for erroneous disparity detection. In the end, an experiment with the dynamic case, when the fixated object is moving, will be discussed.

7.1 Experiment with clear disparity

We simulated the divergence movement from a point close to the head (104 cm) to a point on an object in front of the head (185 cm). During the divergence movement best focus was kept on the point of fixation. Left and right images were acquired five times during this movement uniformly distributed in the range. Between each image pair normalized correlation was performed as a matching criterion in the symmetric manner described earlier and the highest peak was extracted. To this correlation peak is associated a focus value in each image, given by the variance of the grey-levels. The variance should be higher when an object is better focused. In order to determine whether we in a certain image pair are looking at the correct peak there are two criteria we can use:

- **A correct peak should continuously move to the right (or left) as divergence is performed.**

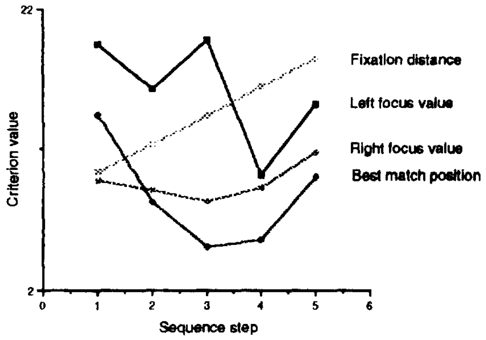

Figure 9: The peak position is the position of the maximum correlation peak. A negative value on the peak position is associated with an object beyond the fixation point (positive in front). The value is half of the disparity.

- If the peak is moving towards the center, the focus values associated with the peak should both increase (decrease if moving outwards).

These criteria are both qualitative and need no knowledge about vergence angles or baseline. If we however know both the vergence angles and the baseline we can also predict how the position, the peak and the focus should change. This proved to be unnecessary in this case.

Figure 9 shows the result of an experiment. As seen in the figure, the first criterion is violated between the first two image pairs. The last three are however consistent. But since the focus value decreased going from the third to the fourth image pair although the peak moves towards the center, the second criterion is violated. The changes are however consistent between the two last image pairs and the conclusion is then that this is a correctly detected peak and that the vergence procedure has succeeded. It is seen from Figure 10 that this is indeed a correct conclusion.

7.2 Experiment involving blur

A point in the left image moving away from the center should become more defocused and a point moving towards the center should become sharper (this should hold also for the right image except with the opposite relationship). This restriction is used to mask off corresponding areas in the correlation image violating this; as seen in Figure 11. In the area surrounding the center, no judgement can be made since points in this area have changed from lying to the left to lying on the right. The first correlation image has no part masked since it is the result of the first image pair and therefore no judgement of focus change can be made. In the following images, the previous mask is superimposed in order to incorporate earlier focus violations. In all but the last image the mask is only effective in the right half since the object is getting better focused. In the last image the position of best focus is passed and the left

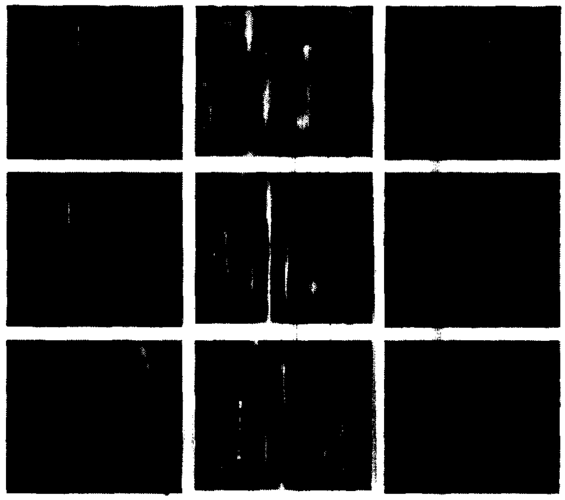

Figure 10: The left and right images from the last three divergence steps are shown together with their correlation image. The white lines in the left and right images are indicators for how successful matchings are and stand for the positions corresponding to the highest peak in the correlation image. The correlated areas in images are depicted by the two parallel lines; the area inside the white lines is the correlation area. As the divergence movement proceeds, the portions of the white lines are gathered around the correct fixation area.

half is also masked off. The band left in the left part is due to new data coming in which have no previous observations.

The sequence in Figure 11 begins from the top and is completed at the time when the third image pairs (from top) are grabbed. The next step of divergence (the image pair at the bottom) is the overshoot that confirms the validity of the fixation at the third image pair by detecting that the sharpness is decreased in the last image pair. This sequence demonstrates also the case when the disparity detection fails in the region defined by depth of field.

7.3 Real-time tracking of a dynamic object

In the vergence experiments shown previously the environment was static. In order to perform the same experiment in a dynamic environment (i.e. fixate on a moving object) it is essential to maintain the symmetry during the vergence movement. This is done by a *parallel* stabilizing process which compensates for asymmetric movements of the object with symmetric version movements. This can be seen in Figure 12, where the same symmetry is maintained while the person is moving in front of the camera. This version process is now totally *independent* of the vergence process since the symmetry is not affected by the symmetric vergence movements. This shows that the vergence strategy described in this paper has the possibility of functioning not only in a static environment but also in a dynamic environment. In fact it seems as if it will function

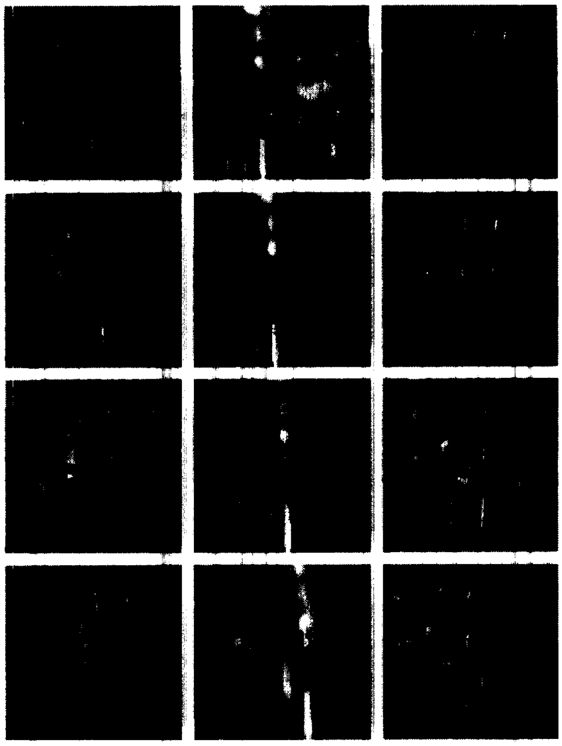

Figure 11: The left and right images from the last four divergence steps are shown together with their correlation image. The variance method is used here as a sharpness criterion.

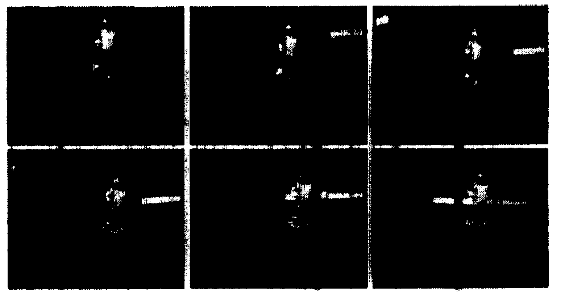

Figure 12: A sequence of dynamic tracking of a moving person in real-time. The movement is at a speed of about 50 degrees per second. Every fifth frame is depicted from top-left to bottom-right. In order to see the symmetries, the fusing vergence movements are disabled.

even better dynamically since only the symmetry of the tracked object is maintained in time while false symmetries will only occur spuriously in time.

References

[1] M.Alpern and P.Ellen, *A quantitative analysis of the horizontal movements of the eyes in the experiment of Johannes Müller*, American Journal of Ophthalmology, 42, 289-303 1956

[2] M.Alpern, *The Position of the Eyes During Prism Vergence*, American Journal of Ophthalmology, 57, 345-353 1957

[3] R.H.S. Carpenter, *Movements of the Eyes*, Pion Press, London, 1988

[4] G.A. Fry, *An Experimental Analysis of the Accommodation-Convergence relationship*, American Journal of Optometry, 14, 402-414 1937

[5] G.A. Fry, *Further Experiments on the Accommodation-Convergence Relationship*, American Journal of Optometry, 16, 325-336 1939

[6] A. Horii, *The Focusing Mechanism in the KTH Head-Eye System*, TRITA-NA-P115, Royal Institute of Technology, Computational Vision and Active Perception Laboratory, Stockholm, 1992

[7] A. Horii, *Depth from Defocusing*, TRITA-NA-P116, Royal Institute of Technology, Computational Vision and Active Perception Laboratory, Stockholm, 1992

[8] B.K.P. Horn, *Robot Vision*, The MIT Press, p 25, 1989

[9] R.A. Jarvis *Focus Optimisation Criteria for Computer Image Processing*,Microscope, 24(2), p 163-180., 1976

[10] B. Julesz, *Foundations of Cyclopean Perception*, University of Chicago Press, 1971

[11] H.A. Knoll, *Pupillary Changes Associated with Accommodation and Convergence*, American Journal of Optometry, 26, 346-357 1949

[12] E.P. Krotkov, *Active computer vision by cooperative focus and stereo*, Springer Verlag, 1989

[13] R.K. Luneburg, *Mathematical Analysis of Binocular Vision*, Princeton University Press, Princeton, 1948

[14] E.E. Maddox, *Investigations in the Relation Between Convergence and the Accommodation of the Eyes*, Journal of Anatomy 20, p. 475-568, 1886

[15] E. Marg and M.W. Morgan, *The Pupillary near reflex. The Relation of Pupillary Diameter to Accommodation and the Various Components of Convergence*, American Journal of Optometry, 26, 183-198 1949

[16] E. Marg and M.W. Morgan, *Further Investigation of The Pupillary near reflex; the Effect of Accommodation, Fusional Convergence and the Proximity Factor on Pupillary Diameter*, American Journal of Optometry, 27, 217-225 1949

[17] J.Müller, *Zur Vergleichenden Physiologie des Gesichtssinnes*, C Conbloch, Leipzig, 1826

[18] K. Pahlavan, T. Uhlin and J.O. Eklundh, *Integrating Primary Ocular Processes*, Proc. 2nd ECCV, Santa Margherita Ligure, Italy, May 1992, pp. 526-541

[19] A. Pentland *A New Sense for Depth of Field*, Proc. IEEE Trans. PAMI, p 523-531. 1987

[20] M. Subbarao *Parallel Depth Recovery by changing Camera Parameters*, Proc. 2nd ICCV, p 149-155. 1988

[21] J.M.Tenenbaum *Accommodation in Computer Vision*,Ph.D thesis, Stanford University, 1970

[22] A. Yarbus, *Eye Movements and Vision*, Plenum Press, New York, 1967

[23] W. Zhang and F. Bergholm, *An Extension of Marr's "Signature" Based Edge Classification*, Technical Report 104, Royal Institute of Technology, Computational Vision and Active Perception Laboratory, Stockholm, 1992