Phase-Based Disparity Estimation in Binocular Tracking

Atsuto Maki*, Tomas Uhlin and Jan-Olof Eklundh

Computational Vision and Active Perception Laboratory (CVAP)
Department of Numerical Analysis and Computing Science
Royal Institute of Technology, S-100 44 Stockholm, Sweden
Email:maki@bion.kth.se, tomas@bion.kth.se and joe@bion.kth.se

May 1993

Abstract: Binocular robots whose cameras can be independently directed require some mechanism for verging; i.e. aiming both cameras at a fixation point. Since the desired vergence angle is directly related to target distance, some sensory cue to give depth information is required to the vergence system. The estimation of binocular disparity is a fundamental precursor to the depth estimation. Algorithms using phase provide a new promising method for example based on the output of Gabor filters for disparity estimation. In this article, at first, we give a brief description of the phase-based algorithm, including an extension to handle the "wrap-around" problem, and a coarse-to-fine strategy "the hierarchical method with pixel-shift" is presented. For binocular tracking, computational time is restricted. Nevertheless, generally heavy calculations are required to get an output through Gabor filters. For that reason, a step-wise fast filter is examined and it is shown to be a substitute for the Gabor filter. Finally the algorithm with the fast filter is applied to a tracking control on the KTH head-eye system.

Keywords: phase-based algorithm, Gabor filter, disparity, vergence, wrap-around, coarse-to-fine strategy, head-eye system, binocular tracking

1 Introduction

Disparity is defined as the angle of correspondence of two associating patterns in the left and the right image, respectively, and many different approaches have been suggested for this measurement, including the two main approaches; correlation-based techniques and correspondence-based techniques [JJT91]. But although many of them give good results they appear less satisfactory for tracking control, which requires both high robustness and low cost "direct" computations. A new method which is promising in this aspect based on the output phase of bandpass Gabor filters has recently been described[San88][BCG90][JJJ91][WWK92]. The objective of this article is to adapt the phase-based approaches to depth reconstruction and to implement the scheme to dynamic tracking control for binocular robot heads. The contents of the article has the following outline.

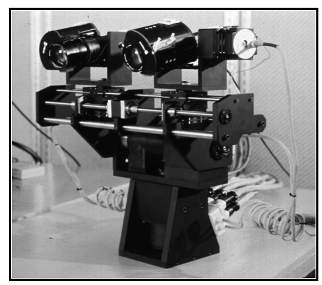


Figure 1: KTH head-eye System

In the next chapter, we briefly describe the general idea of the phase-based algorithm and the Gabor filter. As an extension to handle the "wrap-around" problem, we propose a coarse-to-fine strategy, "the hierarchical method with pixel-shift". A step-wise fast filter is also presented as a substitute for the Gabor filter. After introducing the concept of the binocular vergence in Section 3, the experimental results with the phase-based algorithm are shown in Section 4. There we make a comparison of the results from a Gabor filter and the fast filter, and show the fast filter is competitive and even has an advantage concerned with the normaliza-

*Electrical Engineering Department, University of Tokyo.

tion problem. The algorithm involving the improved methods is finally applied to real-time tracking control on the KTH head-eye system.

2 The Phase-Based Algorithm

The basic concept of the phase-based algorithm is to convolve the left and right stereo images with a complex filter, such as a Gabor filter, and to use the complex phase difference of the filter output. As the local shift between two images is linearly proportional to the local phase difference, the disparity is obtained at each point with an approximation.

2.1 Description of the Algorithm

Noting that the stereo cameras are located on a horizontal axis, we describe the algorithm on the basis of 1-dimensional(horizontal) disparity estimation. At first, $l(x,y)$ and $r(x,y)$ are defined as a stereo image pair from the left and right respectively. And we consider about the case that a constant horizontal disparity (Δx) is found across some part of the scene then in that region we have a relation:

$$l(x,y) = r(x+\Delta x, y). \quad (1)$$

We recover the value Δx from the products which are obtained by convolving $l(x,y)$ and $r(x,y)$ with a complex filter $f(x)$, that is:

$$\begin{aligned} conv_r(x_0,y_0) &= \int r(x,y_0)f(x-x_0)dx \quad (2) \\ conv_l(x_0,y_0) &= \int l(x,y_0)f(x-x_0)dx \\ &= \int r(x+\Delta x,y_0)f(x-x_0)dx. \quad (3) \end{aligned}$$

These products can be calculated at any point (x_0,y_0) and a particular complex filter(see Section 2.2).

Disparity Measurement
Practically the calculations above can also be done in Fourier domain considering the Wiener-Khintchine's theorem. And regarding the integral range in the Fourier domain as the neighborhood of the filter frequency ω_0[radian/pixel], we get a relation in frequency shift between $conv_r(x_0,y_0)$ and $conv_l(x_0,y_0)$:

$$conv_l(x_0,y_0) \approx e^{j\omega_0\Delta x}conv_r(x_0,y_0). \quad (4)$$

Now the disparity Δx is approximated by computing the complex phase difference $\Delta\Phi$:

$$\begin{aligned} \Delta x &\approx \Delta\Phi/\omega_0 \quad (5) \\ \Delta\Phi &\equiv arg[conv_l] - arg[conv_r]. \quad (6) \end{aligned}$$

This is strictly valid only for filters of infinitesimal bandwidth, arising directly from the Fourier shift theorem[San88].

Confidence Value
While the convolution products have some responses at any pixel in the image, the responses in areas where little variance can be seen in the horizontal direction are not reliable. Therefore it is significant to have a confidence value as a threshold to see the reliability of the disparity estimation. To handle this problem, we define a confidence value as follows:

$$conf \equiv \frac{mag[conv_l] \times mag[conv_r]}{mag[conv_l] + mag[conv_r]}. \quad (7)$$

The definition of the confidence value(eq. 7) is based on the magnitude value of the convolution product(eq. 4) since the odd filter gives a response on the variance in horizontal direction(*e.g.*, vertical edge) and the even filter on the vertical line. And the magnitude values from left and right products are multiplied so as to assure that both responses are high enough. At the same time, the multiplied value is divided by the additive product to make a compressing scaling.

2.2 Filters

Gabor Filters
In the phase-based algorithm, it is important that both the spatial width of the filters and the spatial frequency bandwidth are small. Nevertheless, the well-known uncertainty theorem for spatial width and spatial bandwidth dictates that a local filter must have a nonzero bandwidth. One-dimensional filters which minimize the product of spatial width and bandwidth were first described in the time/frequency domain by Gabor [D.46]. And these filters and their Fourier transforms have the functional form:

$$\begin{aligned} g(x-x_0) &= e^{\frac{-(x-x_0)^2}{2\sigma^2}} \cdot e^{j\omega_0(x-x_0)} \quad (8) \\ G(\omega-\omega_0) &= e^{\frac{-(\omega-\omega_0)^2}{2\tau^2}} \cdot e^{-jx_0(\omega-\omega_0)}. \quad (9) \end{aligned}$$

x_0 : The spatial location of the filter.
ω_0 : The central frequency of the power spectrum.
σ : The spatial half-width of the filter.
τ : The half-width in frequency domain($\sigma\tau = 1$).

These "Gabor filters" are thus produced by the multiplication of a Gaussian envelope and complex harmonic function. As for the accuracy of a disparity estimation, it is very sensitive to the spatial frequency of the filter. In our analysis, however, the parameters

$\omega_0 = 1.0$ and $\omega_0\sigma = 1.3$ are found to keep the most reliable performance to detect 1 pixel's disparity.

Fast Filter

For our aim, real-time object tracking, the calculation of disparity estimation should be quick. Though Gabor filters are theoretically efficient, they usually require heavy computations especially for the calculation of the 2-dimensional Fourier transform. Thus, as a substitute for the Gabor filter, we form a pixel-wise complex filter $p(n)$[n:0,1,2,...].

$$Im[p(n-n_0)] = \begin{cases} -t, & \text{for } n = n_0 - 1 \\ t, & \text{for } n = n_0 + 1 \\ 0, & \text{for } n \neq n_0 \pm 1 \end{cases} \quad (10)$$

$$Re[p(n-n_0)] = \begin{cases} -t, & \text{for } n = n_0 \pm 2 \\ 2t, & \text{for } n = n_0 \\ 0, & \text{for } n \neq n_0 \pm 2, n_0 \end{cases} \quad (11)$$

n_0 : The spatial location.

t : The amplitude scale.

The frequency of this filter is regarded as $\omega_0 = \pi/2$[radian/pixel]. And both of the filters are adapted for the algorithm and the results are compared each other(Section 4.2).

2.3 Coarse-to-Fine Strategy

Problem from Wrap-Around

In the practical application of this algorithm, the phase difference $\Delta\Phi$(eq. 6) will be modulated into the range $(-\pi, \pi]$. This means disparity estimates are always expressed within the range $(-\lambda/2, \lambda/2]$ (λ: wavelength of the filter). In other words, the maximum disparity which a filter can accurately determine is one-half the wavelength of filter due to the "wrap-around".

$$\Delta x \leq \frac{\lambda}{2} = \frac{\pi}{\omega_0} \quad (12)$$

For this reason, large filters are necessary to determine large disparities and at the same time these large filters will only give the information about the low frequencies in the image.

The Hierarchical Method with Pixel-Shift

Faced with this problem, a couple of methods have been proposed to keep the performance of the algorithm high, including a coarse-to-fine strategy[JJ89] [WWK92] [CM92]. Our procedure "hierarchical method with pixel-shift" is also based on a coarse-to-fine pyramid and described as follows:

1° Construct a coarse-to-fine Gaussian pyramid. Levels are separated by an octave in scale.

2° Begin to calculate the disparity estimation at the coarsest level. The estimation obtained is taken over to the next level as a shift information.

3° At each level, make a disparity estimation by the phase-based algorithm. The pixel referred to is indicated by the shift information with respect to the disparity estimation obtained at the former level(Fig. 2).

4° Repeat 3° down to the fine level. The disparity estimation at the base level is taken as Δx.

3 Binocular Vergence

The "vergence angle" of a binocular system is the angle between the optic axes of its cameras at fixation. The vergence angle, baseline and gaze direction of a binocular system determine a particular fixation point (Fig. 3). The function of the vergence system is in a narrow sense to control the distance from the cameras to the fixation point near some target object. Thus, the vergence problem can be defined as that of controlling the vergence angle to keep the fixation distance appropriate for the current gaze target [Coo92]. In the case of primates, fixation consists of two separate classes of movements. These movements are vergence and version. As far as the disparity is concerned, the two movements associate with two kinds of disparities. Pure version associates with zero disparity and pure vergence with symmetric disparity [UPE92]. The phase-based algorithm gives information for both of these disparities. Here we consider the condition that the fixation point has a stereoscopic disparity of zero and points nearby tend to have small disparities, regarding the correspondence problem as solved by precategorical processing. In Section 4.3, an vergence experiment on KTH head-eye system [PE92](Fig. 1) is presented.

4 Experiment

The phase-based algorithm involving the improved methods are investigated in this section through the experiments. The hierarchical method with pixel-shift is first examined and then the performance of Gabor filters and the fast filters is compared each other. The experiment for a real-time vergence, the main contribution of this article, is carried out in Section 4.3.

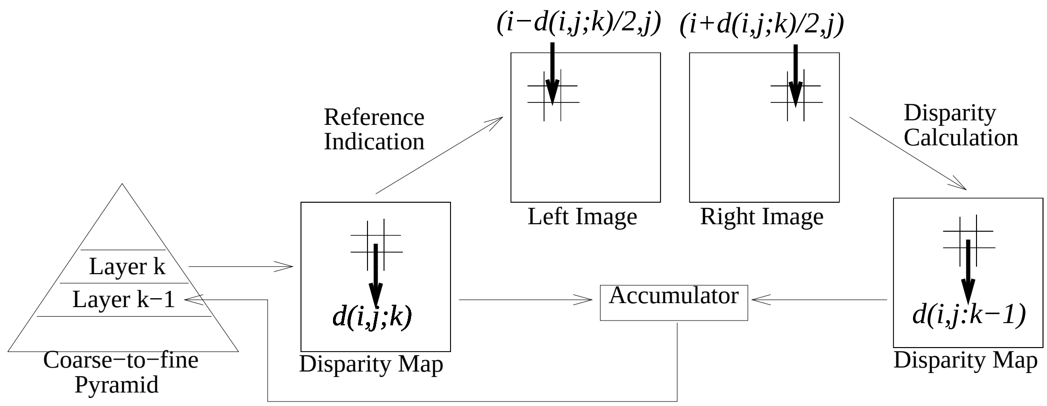


Figure 2: Disparity estimation in the layer k-1.

Layer k: Disparity estimation $d(i,j;k)$ is obtained at pixel(i,j).
Layer k-1: Corresponding to the pixel (i,j), we calculate $d(i,j;k-1)$ by referring $l(i-d(i,j;k)/2,j)$ and $r(i+d(i,j;k)/2,j)$. Additive product of $d(i,j;k-1)$ and $d(i,j;k)$ is the disparity estimation in this layer.

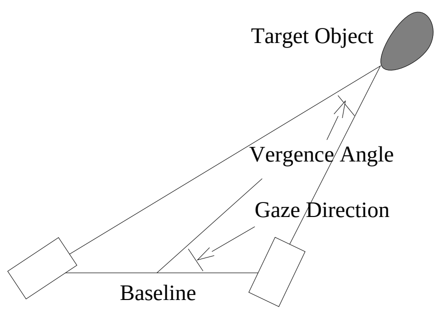
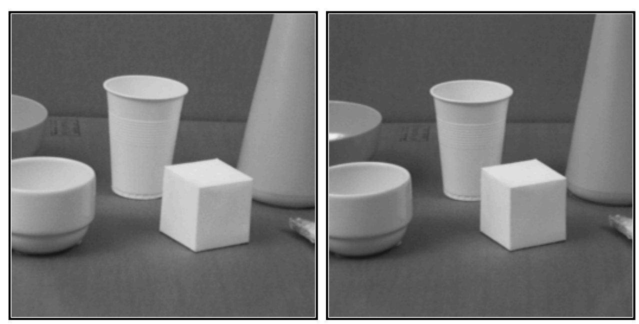


Figure 3: Binocular Gaze Geometry



(a)Left Image (b)Right Image

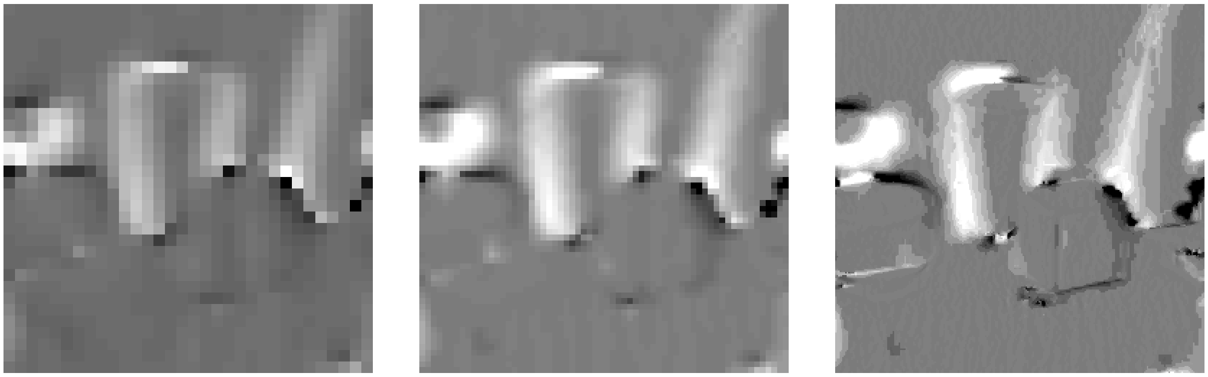
Figure 4: The sample stereo image pair. Both cameras are fixated at the center (left-upper corner of the box) in the each image.

4.1 Hierarchical Method with Pixel-shift

The stereo image pair in Fig.4 is taken up as a sample. The size of the images is 256×256 pixels along x- and y-coordinate. Images are normalized; the average pixel value is set to be 0 and average deviation 1. Several objects are included in the image pair and disparities can be seen in the horizontal and the vertical direction. Our interest is now in the horizontal disparities.

Disparity Estimation
Starting with the highest(coarsest) layer in the pyramid, we have applied the 4-layered hierarchical method with pixel-shift to the image pair. The Gabor filters with the parameters $\omega_0 = 1.0$ and $\omega_0\sigma = 1.3$ are used in each layer. Those filters are expected to cover relatively wide range of disparity as detectable, since those filters with variant scales give also reliable estimations. In our hierarchical method, this fact is especially significant to give good initial estimations.

Fig. 5 shows the disparity map obtained at each layer in the pyramid. Disparities are estimated first roughly then finely. The gray level is corresponding to the disparity(depth) estimation; positions at darker pixels are estimated closer than the fixation point, the center of the image, and positions at lighter pixels further. Note that the reliable estimations are obtained in the regions where vertical(or vertically slanted) edges exist. And error estimations are detected in some regions especially due to the vertical disparities. Those estimations, however, will be eliminated by thresholding



(a)Layer 3(32×32) (b)Layer 2(64×64) (c)Layer 0(256×256)

Figure 5: Disparity Map by Gabor Filters

(a)Layer 3(32×32) (b)Layer 2(64×64) (c)Layer 0(256×256)

Figure 6: Confidence Map by Gabor Filters

with the confidence values.

Thresholding

To see the reliability of the estimations, we calculate the confidence values as well. Fig. 6 shows the confidence map obtained at each layer. Higher(lighter) gray level at a pixel is corresponding to higher confidence value. The confidence values obtained in one layer are accumulated to the ones in the next layer. High confidence values are observed where vertical edges exist in the original images. Referring the confidence map, the disparity map in Fig.5(c) is thresholded [1] and the result is shown in Fig.8(a). The precision of the disparity estimations is up to 0.5 pixels in this final stage. In Fig.8(a), it should be also noted that the disparity is not properly detected at the vertical edge in the center of the box(see Fig.4). This is because the pixel values across the edge are on both sides kept positive even after the normalization. In the case images include edges in various gray levels, iterative normalizations and disparity calculations may become necessary to complete the whole disparity map.

[1]The disparity map in Fig.8 is obtained, using a threshold value investigated empirically.

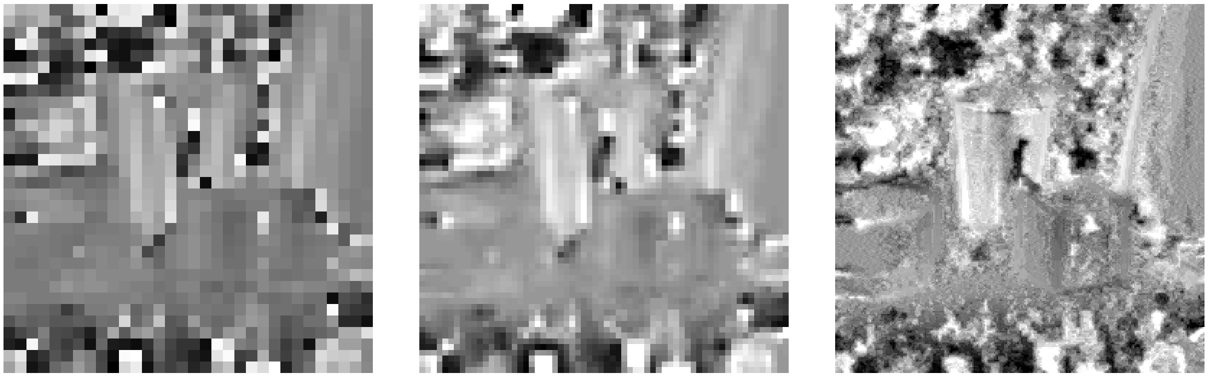
4.2 Gabor Filters and the Fast Filter

We have seen the performance of our coarse-to-fine strategy, the hierarchical method with pixel shift. In this section, the fast filters, as well as the Gabor filters, are applied to the algorithm and the results are compared.

Performance of the Fast Filter

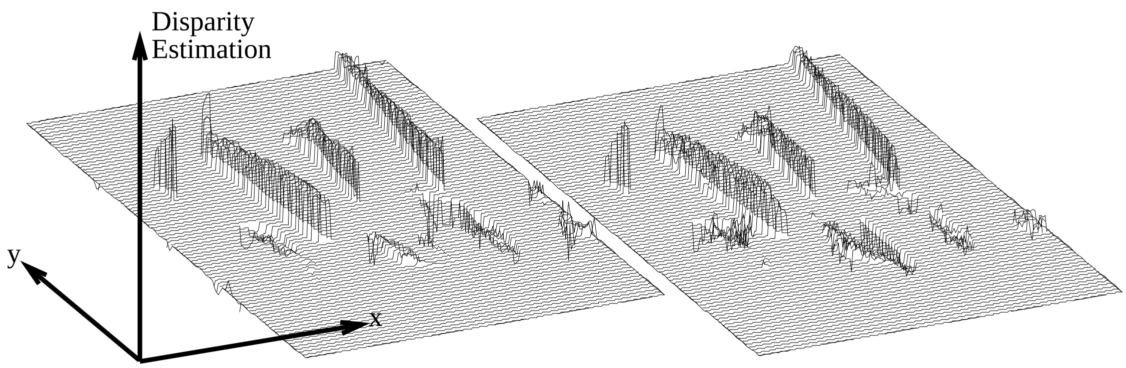
Replacing the Gabor filters with the fast filters, the hierarchical method is adapted in the same manner as in the former section, but without normalizing the sample images. The disparity estimation is shown in Fig.7, corresponding to the result in Fig.5 obtained with Gabor filters.

In the disparity map in Fig.7 noticeable noise is present. This is because both of the odd and even part of the fast filters have zero average in the outputs. That is, discontinuities occur in the phase where real outputs have infinitesimal values while the imaginary outputs have positive and negative values mutually, and this results in noise in the disparity estimations.



(a)Layer 3(32×32) (b)Layer 2(64×64) (c)Layer 0(256×256)

Figure 7: Disparity Map by Fast Filters



(a)By Gabor filters (b)By Fast filters

Figure 8: 3-D Thresholded Disparity Map

The noise is, nevertheless, eliminated by thresholding with the confidence values.

Thresholded disparity estimations are shown in Fig.8(b). As for the precision, it is not completely equivalent to that from the Gabor filters(compare with Fig.8(a)) but still up to 1 pixel. On the other hand, the disparity at the vertical edge in the center of the box(see Fig.4) is successfully detected. This is what the Gabor filters failed to estimate properly due to the normalization problem. Here we can see the advantage of the "normalization-free" filters.

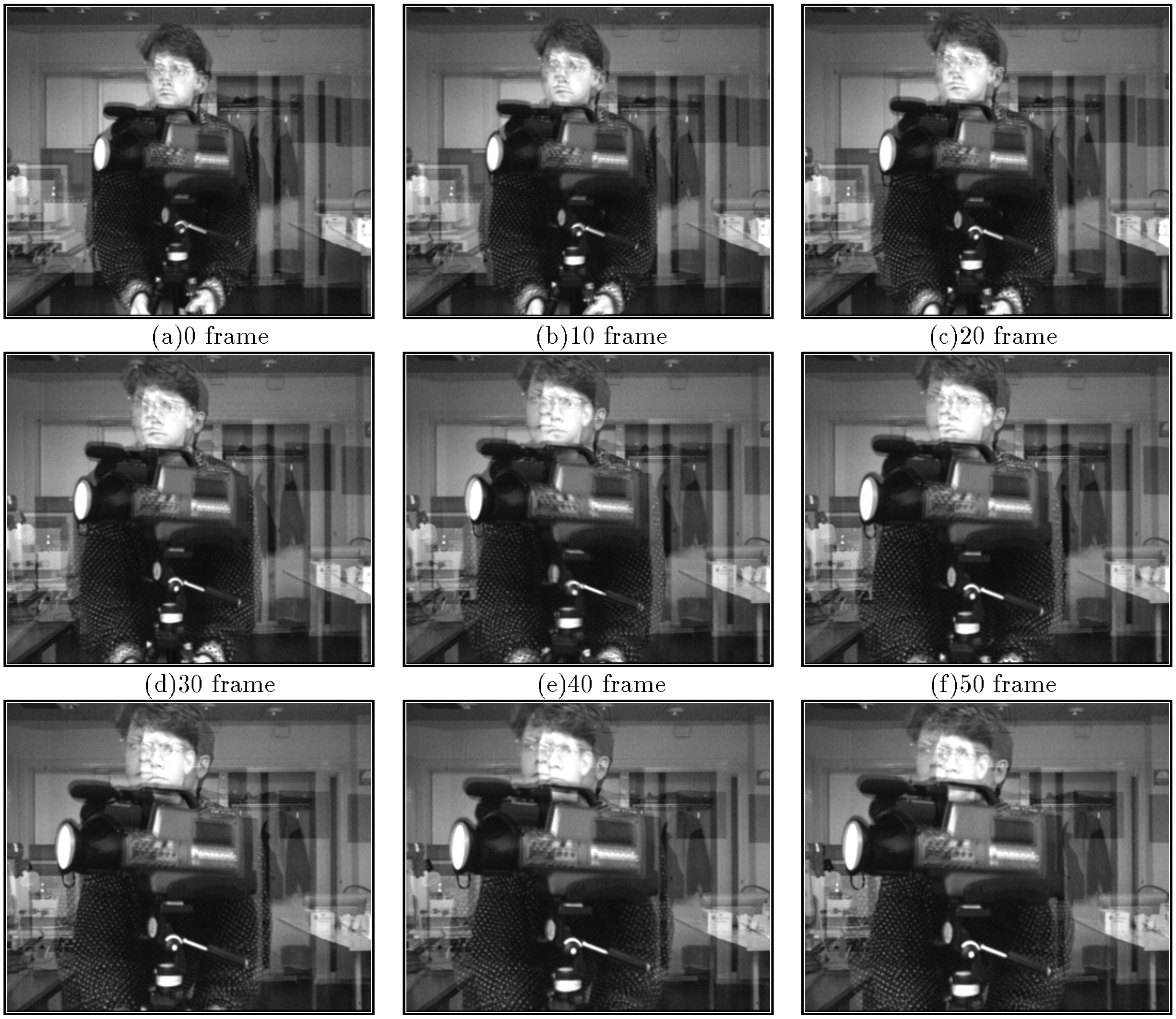
Computational Time

Another big advantage of the fast filters is literally the fast calculation. Table.1 shows the computational time to get the disparity estimation for each of Gabor filters(Fig.5) and the fast filters(Fig.7). The computational cost for the normalization process, which is necessary in the usage of the Gabor filters, is relatively small. Yet, considering the disparity estimation itself, the cost with Gabor filters is nearly twice as much as that with the fast filters. The dominant factor which causes the difference exists in the convolution process. In the case Gabor filters are used, heavy computations for the calculation of the 2-dimensional Fourier transform are required, while they are not involved in the case of the fast filters. Taking the advantage of the normalization process into account as well, the fast filters are chosen in the application to binocular vergence in the next section.

Table 1: Computational Time

	Gabor filters	Fast filters
Normalization	0.13	-
Disparity Estimation	16.06	8.29
Total	16.19	8.29

(CPU Time[sec]/SUN Spark Station 10 Model 20)



(a)0 frame (b)10 frame (c)20 frame
(d)30 frame (e)40 frame (f)50 frame
(g)60 frame (h)70 frame (i)80 frame

Figure 9: Binocular Tracking; images from left and right are superimposed.(1 frame = 40 msec)

4.3 Binocular Tracking

For real-time vergence, the calculation of the disparity estimation is desired to be as fast as possible. From this point of the view, the phase-based algorithm has been implemented on a parallel computer and applied to binocular vergence control.

In order to use the disparity estimate to control the motors, the corresponding angle has to be computed. The number of pixels shift in the image are transformed into angle through the focal length (assuming rotation about the optical center) with the simple relationship

$$angle = tan^{-1} \frac{shift}{focallength}, \tag{13}$$

where the focal length is given in pixels. The correction for disparity is equally divided between the two eyes giving the angular correction to the tracker for each eye of

$$angularcorrection = tan^{-1} \frac{disparity}{2 \times focallength}. \tag{14}$$

This angle will be corrected by the tracker. Thus the object is tracked while it is moving in depth by continuously updating the measurement of disparity. A predictor is used by the tracker to avoid that the time lag from the image processing and control, results in constantly lying behind the tracked object.

Fig. 9 shows how the stereo cameras track the object, here a video camera held by a man. The initial fixation point is at the center of the image. Disparity is calculated around the fixation point; for the region

50×50 pixels around the center. The calculation is performed in every frame(1 frame = 40 msec). The vergence fixation is kept at the center in the images while the object is moving closer. The pictures are taken every 10 frame and those from left and right cameras are superimposed. As the fixation point moves closer, the background gets blurred and there the convergence movement is implied.

In Fig. 10, the pan movement of each camera is plotted. A symmetrical movement can be seen in the left and right pan angles. As a whole, the result shows that the system combining the disparity estimation scheme with the vergence control is working properly.

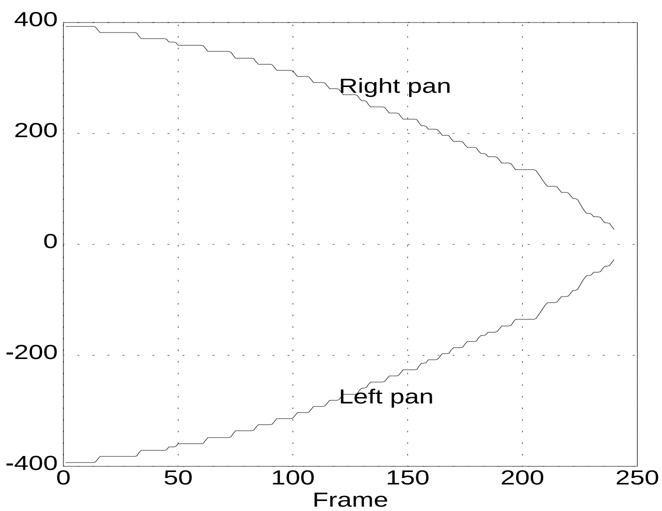


Figure 10: Pan movement of both cameras

5 Summary and Discussion

As stated in the introductory chapter, our aim has been to control the vergence mechanism of the KTH-head. For this purpose, we have considered the essential problem to develop a robust and fast algorithm for disparity estimation and proposed improved methods for the phase-based algorithm:

- a hierarchical method with pixel-shift.
- a fast filter as a substitute for Gabor filter.

The phase-based algorithm and the proposed schemes are investigated through the experiments and the following factors are admitted:

- High precision of the disparity estimation by the hierarchical method.
- The advantage of the fast filter concerning normalization.

Further, those methods have been applied to dynamic vergence control. Through the experiment in cooperation with the tracking system, the whole scheme turned out to work properly in real-time vergence.

Future efforts could take several directions, including:

- Investigations for more sophisticated bandpass filters.
- Hierarchical methods for vertical disparity detection.
- Robust methods in thresholding disparity.

Solutions for those problems will contribute to more robust and integrated vergence and tracking control.

References

[BCG90] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. PAMI*, 12(1):55–73, January 1990.

[CM92] Li-Dong Cai and John Mayhew. Estimating mean disparity of stereo images using shift-trials of phase differences. *BMVC Proceedings*, pages 237–246, September 1992.

[Coo92] David John Coombs. Real-time gaze holding in binocular robot vision. *Dissertation, University of Rochester*, June 1992.

[D.46] Gabor D. Theory of communication. *JIEE*, 93:429–459, 1946.

[JJ89] Allan D. Jepson and Michael R. M. Jenkin. The fast computation of disparity from phase differences. *IEEE CVPR Proceedings*, pages 398–403, June 1989.

[JJJ91] Michael R. M. Jenkin, Allen D. Jepson, and David J.Fleet. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210, March 1991.

[JJT91] Michael R. M. Jenkin, Allen D. Jepson, and John K. Tsotsos. Techniques for disparity measurement. *CVGIP: Image Understanding*, 53(1):14–30, January 1991.

[PE92] K. Pahlavan and J.O. Eklundh. A head-eye system - analysis and design. *CVGIP:Image Understanding*, 56(1):41–56, July 1992.

[San88] T. D. Sanger. Stereo disparity computation using gabor filters. *Biol. Cybern.*, 59:405–418, 1988.

[UPE92] T. Uhlin, K. Pahlavan, and J.O. Eklundh. Fixation by active accommodation. *Proc. SPIE Symp. on Intelligent Robotics and Computer Vision*, November 1992.

[WWK92] J. Wiklund, C-J Westelius, and H. Knutsson. Hiararchical phase based disparity estimation. *Report, Computer Vision Laboratory, Linköping University, Sweden*, (LiTH-ISY-I-1327), February 1992.