Real-time Vergence Control using Local Phase Differences

Michael Hansen, Gerald Sommer.

Computer Science Institute, Christian-Albrechts-University Kiel

gs,mha@informatik.uni-kiel.d400.de

Abstract

We address the control of vergence movement of an active camera system and the computation of disparity maps in real-time. We apply a phase-based approach to estimate horizontal disparity. Besides we use the disparity value of the image center as an error signal for a PD-controller, which has been implemented for movement control because of his stability in real-time systems. To estimate large disparities with small Gabor filters a coarse-to-fine strategy is coupled with the PD-control. The closed-loop vergence control has a performance of 25 Hz. The settling time for convergence on a fixated point is 0.3-0.5 seconds. Due to the phase-based approach sub-pixel accuracy in disparity estimation is possible.

Key words: active vision, stereo disparity, vergence movement

1 Introduction

Depth perception is one of the fundamental problems in computer vision. In the last 30 years a lot of research on algorithms estimating depth from images have been done. Some algorithms based on a single camera, e.q. 'depth from focussing/motion', but many more algorithms use stereo configurations similar to the human eye configuration for depth perception. At first static camera systems were used to recover 3D-information from images. With exact calibration good results could be obtained. A stable solution of the correspondence problem was the key for the perception of depth.

In the active vision paradigm in the sense of Bajcsy [4] cameras become an active instead of passive observer. The acquisition of visual data is no longer a passive process, but an active one. It needs to be controlled by optimal strategies depending on the goal of the vision process. Ballard and Brown [11] called this *animate* or *purposive* vision. They involved the principles of human vision like learning and gaze-control in computer vision. In the human visual system the gaze control is a means to select necessary information for visual perception. Therefore control of the eye movements has to be seen as a part of the system. Action is related with perception in the so called action-perception cycle (Sommer [17]). For example the vergence movement of the eyes are related with depth perception in the near field of visual perception.

The reasons for vergence movement in an artificial vision are not so obvious as for the human visual system with its nonuniform spatial resolution of the retina. Normally common technical sensors have a uniform but not retina like topology of pixels. Studying the cooperation between vergence movement and the different eye (camera) movements as smooth pursuit and saccades, described by Yarbus [2] for human vision, is one aspect of basic research in active vision. Some more practical aspects for verging are based on mathematical reasons, making analysis of objects near the optical axis of both cameras easier. Verging allows disparity-based segmentation, ignoring features with too large disparities in the interpreted scene. These two aspects for vergence movement, noticed from Olsen & Coombs [10], are complemented from the important property to reduce stereo disparity in the scene. Reducing disparity increases the performance of most stereo algorithms for depth estimation. Thus vergence movement is a necessary capability for artificial visual systems. Working with small disparities allows the use of smaller filters and reduces the computation time. We will show in this work that real-time depth estimation at 25 Hz becomes possible.

The angle between the two optical axes of the cameras is called *vergence angle*. It is directly related to the distance of the object which is fixated by both cameras. Thus any depth cues like motion, texture or shading can be used to control this angle. The most common depth cue for

vergence control is horizontal stereo disparity. Converging on a point means that the optical axes intersect at that point, resulting in a zero-disparity there. Our system fixates interesting objects with a dominating (right) camera. The vergence movement is controlled by reducing the measured horizontal disparity to zero.

For real-time applications at 25 Hz we need a fast stereo algorithm using the given pipeline architecture of the image processing system in an optimal manner. Feature based algorithms which solve the correspondence problem explicitly by a search process are unsuitable because of several principal and practical reasons. Region based algorithms like correlation and phase-based approaches give denser disparity information than feature based algorithms, which cover such information only at the features. Correlation needs a maximum search and due to different lighting and contrast conditions in both cameras, correlation is sensitive to noise. An algorithm based on bandpass filters could give more stable results under real world conditions. Gabor filters [1] with limited spatial width and finite bandwidth are suitable filters for computing stereo disparity [5] in such an approach.

We use a phase-based approach with spatial Gabor filters to compute multi-resolution disparity maps in real-time. The disparity in the center of the map is used to control the vergence movement. A coarse-to-fine strategy helps us to deal with large disparity values unmeasurable by small filters, we have to use to obtain real-time performance.

We describe in section 2 related works to control vergence movements using different approaches for disparity estimation. Section 3 explains our way to estimate horizontal disparity with Gabor filters in a multi-resolution manner. The closed-loop control of vergence movement and the system architecture are subject of section 4. In the last section 5 we show results and experiments in disparity estimation.

2 Related Works

Like saccadic eye movements and smooth pursuit the vergence movement is one of the basic capabilities of an active vision system. Thus most research groups working in the field of active vision are implementing modules to control vergence movement. First Abbot & Ahuja [3] use vergence movement for better surface reconstruction. Later Krotkov [6] combine different stereo algorithms with verging cameras for increasing depth perception. He increases the stability of the system by implementing a 'depth from focus' algorithm.

By using a cepstral filter for disparity estimation Olsen & Coombs [10] develop a real-time vergence control with a comparable servo rate of 10 Hz. They use also a PD-controller for vergence control as we do. Due to the characteristics of cepstral filtering they have to handle with multiple peaks in their filter responses. The hardware configuration is similar to our system using a pipeline processor. More closer to our approach in disparity estimation are the work from Theimer & Mallot [15]. They also use a phase-based approach with Gabor filters on sub-sampled images with a rate of 1 Hz on common hardware. They optimized the resulting disparity map by applying the *local frequency model* and different filter orientations. This reduced the servo rate to 0.1 Hz.

Westelius et al. [16] develop a vergence control based on phase differences. In a coarse-to-fine strategy they use a local image shift to deal with large disparities. To get stable results they additionally compute the disparity from a pair of edge images in each resolution and weighted both disparities from image and edge image pair. Their algorithm is implemented on an active system but up to now no performance data were given. The group of Eklundh et al. [18] controlles vergence movement with the same servo rate of 25 Hz as we do. But different to us they use smaller filters in only one dimension and different confidence measures. Their vergence control is integrated in a smooth pursuit system. This will be realized also in our system in the next time.

3 Phase-based Disparity Estimation

The idea of a phase-based approach for disparity estimation is to solve the correspondence problem implicitly. Without explicit feature extraction and search processes phase-based approaches can be described similarly by a local correlation of bandpass-filtered images. The local phase response of the signal contains the information of the spatial position of the matched structure. According the Fourier shift-theorem

$$f(x) \circ\!\!-\!\!\bullet\, F(\omega) \qquad f(x+D) \circ\!\!-\!\!\bullet\, F(\omega)e^{ikD}, \tag{1}$$

a global spatial shift D of a signal $f(x)$ can be detected as a phase shift in the Fourier spectrum. To recover the local disparity as a spatial shift the local phase differences have to be computed. Extracting the local phase in both images of a stereo pair with complex filters like Gabor filters leads to a direct computation of local disparity.

Sanger [5], Langley et al. [8] and Fleet et al. [9] have employed phase-based approaches in one or two dimensions to recover disparity information with complex Gabor filters (2) on different scales.

$$G(x;\sigma,\omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} e^{-i\omega x} \tag{2}$$

We denote the Gabor filter responses $f_g(x) = G(x;\sigma,\omega) * f(x)$. The magnitude of the right image response is $r(x) = |f_g(x)|$ and that of the left image is $l(x)$. The phases are denoted $\phi_r(x) = arg[f_g(x)]$ for the right and $\phi_l(x)$ for the left image.

To have similar filters on different scales the mean frequency ω is linked with the width σ of the Gaussian by the bandwidth factor $t = \frac{1}{\sigma\omega} \in (0,1)$. A bandwidth factor $t = 0.33$ leads to a full bandwidth of one octave.

The spatial shift $D(x) = \frac{\Delta\Phi(x)}{\omega}$ in a stereo image pair is computed from local phase difference:

$$\Delta\Phi(x) = \phi_l(x) - \phi_r(x). \tag{3}$$

This is called the *constant frequency model* $\omega(x) = \omega$. In some cases the phase can change very quickly, not linear as expected. The reasons are singularities in the phase signal (see [9] for details). To overcome this trouble the *local frequency model* $\bar{\omega}(x)$ as the mean of first derivatives with respect to x : $\phi_l'(x), \phi_r'(x)$ of the phases $\phi_l(x), \phi_r(x)$ is defined by Fleet:

$$\bar{\omega}(x) = \frac{1}{2}(\phi_l'(x) + \phi_r'(x)) \tag{4}$$

The local frequency model increases the accuracy of disparity estimation, but also increases the computational costs because of the wrap around of phase from $+\pi$ to $-\pi$ which allows no direct computation of the local phase derivatives (see [9] and [15]).

The linear behaviour of the phases depends on the magnitude of the Gabor filter response. Independent of a weak magnitude in both or one image you can always compute a local phase difference. But a weak magnitude causes a false match to the structure in the images. So all authors have defined different *confidence values* $c(x)$ based on the magnitude $l(x), r(x)$ of the Gabor filter responses in the left and right image to get a measure for stability of the disparity. Some authors ([5],[8],[15]) weighted the estimated disparity of different scales and orientations by their confidence value to get more stable results.

3.1 Our Approach for a fast real-time algorithm

For disparity estimation the choice of the filter parameters are influenced from the given hardware to obtain real-time performance. Small filters and a simple algorithm can perform a high clock rate. We developed such a simple algorithm basing on the theoretical principles of the phase-based approach.

3.1.1 Filter design

By designing filters for measuring phase disparities the following constraints have to be regarded [16]:

a. *No DC component* because a small signal fluctuation compared to the DC level results in a non optimal phase behaviour.

b. *No wrap around* of the phase in the impulse response in order to maximization of the measurable disparity.

c. *Monotonous phase* to assure the one to one relation between phase difference and disparity.

d. *Small spatial support* to get low computational costs.

e. *Only positive frequencies* because of quadrature requirements.

We choose a Gabor filter with an odd size of 7x7 using the maximal possible template size of 8x8 of the given image processing hardware (MaxVideo 200) optimally. A bandwidth factor $t = 0.33$ is chosen to reach a bandwidth of one octave. To get no wrap around and to have a maximum measurable horizontal disparity related to the filter size a wavelength $\lambda = 2\pi/\omega_x = 6$ pixel is optimally.

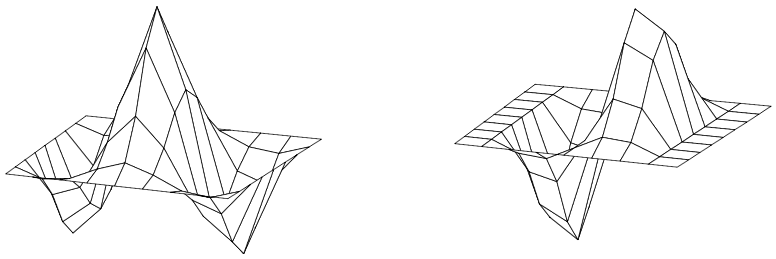

Fig.1: Odd and Even Gabor filter 7x7 with $\lambda = 6$ pixel and aspect ratio $\sigma_x/\sigma_y = 2$.

Like other authors [15] we choose an aspect ratio of the Gaussian $\sigma_x/\sigma_y = 2$ optimizing to vertical edge responses. Fig.1 shows our discrete even (DC corrected) and odd Gabor filter. The filter has to be DC-corrected to get well phase behaviour. In the continuous case the corrected Gabor filter by (5) has no linear phase response [13].

$$G(x;\sigma,\omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}(e^{-i\omega x} - e^{-c^2/2}). \tag{5}$$

The phase-behaviour of our discrete DC-corrected 7x7 Gabor filter shows Fig. 2. Due to a large scale factor $c > 3$ the phase is still linear. So the DC-corrected filter is suitable for phase estimation.

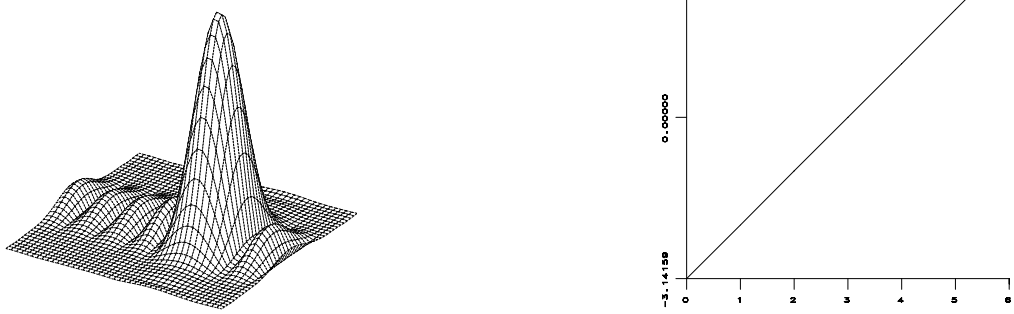

Fig.2: The Fourier spectrum and phase behaviour of our discrete DC-corrected Gabor filter.

The Fourier spectrum of the Gabor filter has a nonzero magnitude in the negative frequency domain due to quantization and cut-off effects. Thus the quadrature requirements are not completely fulfilled. This can cause a change in the sign of phase differences depending on the frequency content of the image [16]. In practice the measured phase differences are linked with confidence values, so this effect could be neglected.

3.1.2 Confidence value

Without a consistency check of the measured phases in the left and the right image disparity estimation can produce arbitrary results. We choose the following way to check the stability of phase information. First in each image the magnitude $r(x), l(x)$ of the filter responses is thresholded (by 20 % of the maximum magnitude). Magnitudes below this threshold are set to zero. Second the sum of magnitudes $l(x) + r(x)$ of both image are calculated and thresholded again (by 40 % of the sum maximum magnitude). The estimated phase difference at image pixel x is called stable if these two constraints are complied. This binary map of stable phase differences is our *confidence map*. The results of a disparity estimation without thresholding are explained in Fig.3: An one dimensional bar (width 4 pixel, height 10 pixel) is shifted with $D = 2$ pixel in the other signal. In the left and middle figure you can see the same behaviour of phase and magnitude of the filter responses in both signals only shifted with D. The right figure shows the estimated disparity (lower curve). It is more exact when the sum of magnitudes is high (dotted curve). Larger disparity reduced the spatial area of reliable disparity estimation due to the small overlapping of the filter responses.

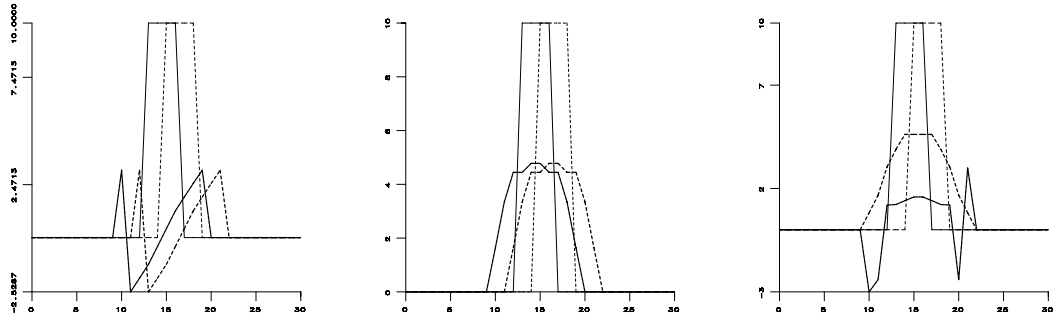

Fig.3 Foreground: (left) Behaviour of phase, (mid.) magnitude, (right) magnitude sum (dotted) and estimated disparity at an one dim. bar. In the backround both signals are overlapped with a shift $D = 2$.

The stable estimated disparity values are between 1.3 -1.6 pixel. The true disparity is 2 pixel. This results are in same range published by other authors [12]. Due to the filter design the results become more exactly if the object stimuli are more sine-wave shaped with the same mean frequency as the Gabor filter.

3.2 Multi-resolution disparity estimation

The small filter size, necessary for real-time performance, demanded a strategy to deal with larger disparities in stereo images. Additionally we have to choose an algorithm which is suitable for real-time processing.

Namuduri et al. [14] presented an approach to compute Gabor filter responses at multiple resolutions. Starting with a high resolution filter response they compute the response at a more coarse level by low-pass filtering with a Gaussian of the same variance as the Gabor filter.

Our approach is to compute a Gaussian pyramid $f_i(x)$ by approximating the Gaussian filter by a 7x7 binomial filter (Low). Sub-sampling (Sub) reduced the image resolution from 512x480 at the finest level to 32x30 at the coarsest level. The Gabor filter response $f_g(x)_i$ at the level i is computed by

$$f_g(x)_i = G(x, \sigma, \omega) * f_i(x) \quad \text{with} \quad f_i(x) = Sub * (Low * f_{i-1}(x)). \tag{6}$$

This results in a maximal measurable disparity of ± 96 pixel at the highest level. Due to the results of the confidence check (3.1.2) the maximal stable disparity is quite smaller. Figure 4-5 show the responses at different resolution levels. The phase response is shown on a scan-line (row 128) in the middle of the poster on the wall. A threshold of 20% of the maximal magnitude are chosen to check the reliability of the phase response.

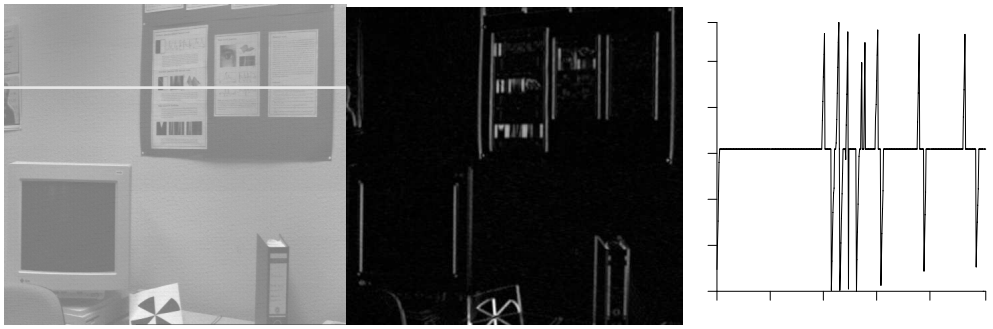

Fig.4: (l.) Responses of Level 1 (512x480) image, (m.) magnitude, (r.) phase (scan-line at row 128, only reliable values).

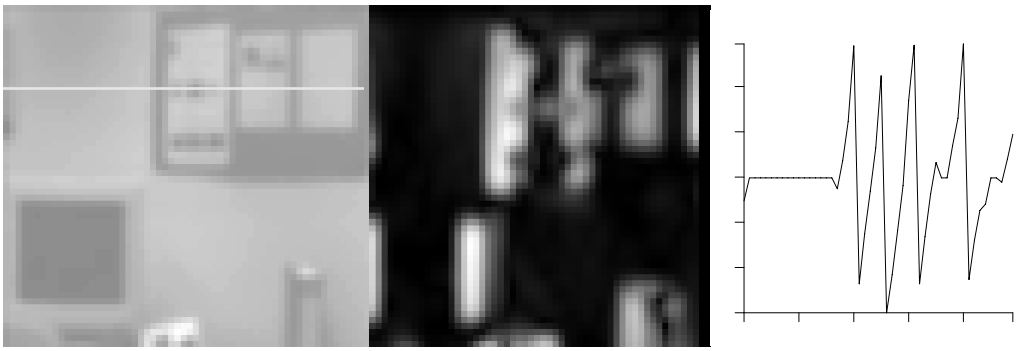

Fig.5:(l.) Responses of Level 4 (64x60) image, (m.) magnitude, (r.) phase.

4 Vergence Control

The Kiel camera mount [1] has four mechanical degrees of freedom: the pan angle of the neck, the tilt angle, and two vergence angle θ_r and θ_l. The right camera has been declared as the dominating eye of our vision system. Pan, tilt and right vergence angle θ_r are controlled by a superior gaze-controller. The object, which should be fixated by both cameras, is first located at the center of the right field of view. Regarding vergence movement we only have to control the left vergence angle θ_l to reduce the horizontal stereo disparity D_c in the center of view. (In a calibrated system no vertical disparity can occur in the center.)

The disparity D_c is picked out from the center of the computed disparity map. To get more robust values of D_c, it is possible to pick the median of a small neighbourhood (3x3) depending on the Gabor filter size. The estimated disparity at a fixated rectangle in the image are only reliable in the small overlap of the Gabor filter responses, so the median of the center area would be a good compromise in a real-time application.

Figure 7 shows the vergence control in a diagram. The PD-controller is denoted by G. In the plant H the motor control and the vision process are modelled.

[1] Consisting of the TRC BiSight Vergence Head and the TRC UniSight Pan/Tilt Base

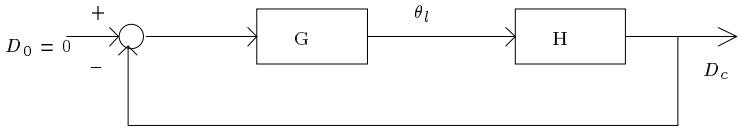

Fig 7. Block diagram of the vergence controller.

The vergence controller is designed as feedback loop. The estimated disparities D_c are compared with the reference signal D_0, which is set to zero due to the requirement $D_c = 0$ in the case of convergence. So D_c is the error signal of the closed loop. The left vergence angle θ_l is controlled by a PD-controller because of the robust behaviour in real-time applications. offset $\Delta\theta_l$ to the actual left vergence angle to compute the new angle θ_l. The offsets $\Delta\theta_l$ are defined by the PD-control law:

$$\Delta\theta_l = K_p D_c + K_d \dot{D}_c \tag{7}$$

The controller gains K_p and K_d are tuned in a simulation by the Ziegler-Nichols method [7] to have robust control and minimal settling time. The behaviour of the PD-control is demonstrated in a simulation in Figure 8: A true disparity of 100 pixel is measured correctly at the first time step. In the ideal case during the following steps the slowly reducing disparity D_c is measured exactly by the vision system. The PD-control reduced the estimated disparity near to zero within 10 cycles (0.4 s). In the noisy case the measurement of the disparity D_c are overlayed with a random error of maximal 33 % [2] of the true disparity. Even with this inaccuracy the PD-control is able reduce the disparity to zero nearly in the same settling time.

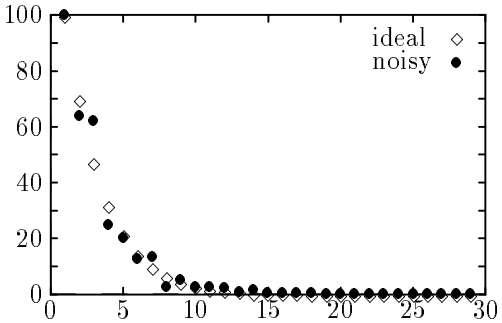

Fig 8. Step responses (noisy and ideal) of a PD-controller ($K_p = 0.3, K_d = 0.05$).

4.1 Vergence control as a closed loop

The whole control loop for vergence movement is performed in a time $t < 40$ ms depending on the resolution of the disparity map. Not until the dominating right camera fixates the object, the lowest resolution (32x30) of the map is computed due to the unknown disparity. If a larger disparity than ± 96 pixels occurs other strategies must be used like 'vergence from accommodation', which can give good results in the nearer field of view [6].

The closed-loop vergence control consists of following steps

1. The analog video signal from the camera are digitized at the image processor in a resolution of 512x480 pixel.
2. The low-pass filtered images are computed by using a binomial filter (7x7) and subsampling until the stated resolution level i is reached.
3. The Gabor responses are computed by convolving with the even and the odd Gabor filter (7x7).
4. The magnitude and phase are computed of the responses with a LUT-Operation with implicit thresholding the magnitude.

[2] The absolute error of 33 % are chosen by Sanger [5] for a linear disparity model due the error estimation for Gabor filters of a bandwidth of one octave ($t = 0.33$).

5. The sum of magnitudes from both images are computed and thresholded giving a confidence map of reliable areas of phase difference.
6. The phase responses are subtracted. Due to the 8 bit arithmetic it is implicit sure that the phase differences $\Delta\Phi(x) = \phi_l(x) - \phi_r(x)$ are between $-\pi$ and $+\pi$.
7. The confidence map is multiplied with the phase difference map.[3]
8. The center of the resulting map D_c is picked out as the error signal for the PD-controller.
9. The offset $\Delta\theta_l$ is determined by the PD-controller.
10. The new vergence angle θ_l is commanded to the motion controller.
11. The motion controller runs the left vergence axis with a control rate of 2 kHz.
12. If D_c was smaller than the half maximal measurable disparity at the next finer resolution the resolution level i is decreased for the next cycles.

The steps 1 to 4 are performed parallel on two MaxVideo 200 boards.The steps 5-7 are computed parallel on both boards crossing the data from one board to the other. The time for image processing (steps 1 -7) are between 33 and 37 ms. The following steps (8-11) are processed at the SUN-Workstation in 3 ms. So the cycle works in servo rate of 40 ms.

5 Experiments

In this section we will show some disparity maps at different resolutions. The example (Fig. 9) is a lab scene with our manipulator in the foreground. The image is grabbed at the end of the vergence movement. The dominating right camera fixates the middle of the manipulator.

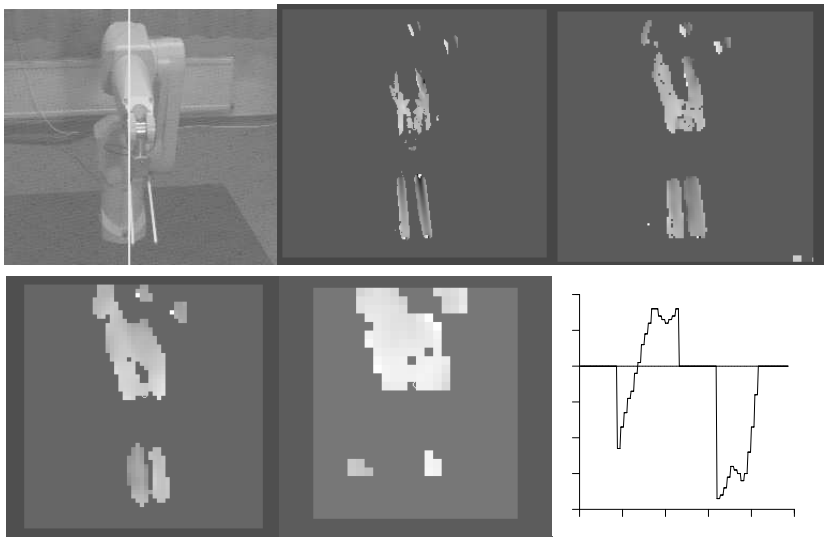

Fig.9:(u.l.) The manipulator of our lab. (512x480), (u.m.) disp. map at level 2 (256x240), (u.r.) at level 3 (128x120), (l.l.) at level 4 (64x60), (l.m.) at level 5 (32x30), (l.r.) behaviour of the disparity (level 4) at the white scan-line (col 116, u.l.).

On each resolution you will get a relative disparity map giving depth information for a local neighbourhood. Figure 9. (l.r.) shows a vertical scan-line (top-down) of the map disparity on level 4. The zero disparity is represented by a constant line. Unreliable disparity values are set to the zero line. The first section of the scan-line resulted from top of the manipulator fixated

[3] Due to the constant frequency model the disparity $D(x)$ is computed from $\Delta\Phi(x)$ only by scaling with the constant mean frequency ω of the filter.

by the right camera. The responses of the fingers are the second part of the scan-line. At the end of the scan-line the smallest disparity is estimated at the finger of the manipulator.

The disparity maps on different scales show some property of the phase-based approach. The spatial bandwidth σ of the Gabor filter are coupled with the mean frequency ω by the bandwidth factor $t = \frac{1}{\omega\sigma}$. The filter responses on different scales depend on the size of structure in the image. First only the vertical edge and the small fingers of the manipulator give reliable responses at level 2. Later at level 5 only the body of the manipulator is matched by the Gabor filter, resulting in reliable disparity values. The disparity values of one resolution can be used to improve the results of another resolution.

Summary

We have shown in this work that vergence control and computing of disparity maps are possible with a clock rate of 25 Hz by using a phase based approach. To increase the accuracy of disparity estimation suitable *confidence values* have to be defined. Later we want to proof if any kind of *local frequency model* is suitable for pipeline processing. By combining the disparity estimation of different resolutions we hope to get depth map computing in realtime in the future. These will be the first steps towards an active vision system for depth perception.

Acknowledgments

Without the helpful support of Gerd Diesner, Henrik Schmidt and Christian Krauss by programming the head system and the pipeline processor this work would not be realized soon. A great thanks to Dr. Kostas Daniilidis for the discussion and the improvement of the presentation.

References

[1] Gabor D. :Theory of communication. JIEE, 1946, 429-459.

[2] Yarbus L.A.: Eye Movements and Vision Verlag. Plenum Press, New York, 1967.

[3] Abott L.A., Ahuja N.: Surface reconstruction by dynamic integration of focus, camera vergence, and stereo. Proc. 2nd Intern. Conf. Comp.Vision, Tampa.

[4] Bajcsy R.: Active Perception. Proceedings of the IEEE, Vol.76, No.8,1988, 996-1005.

[5] Sanger T.D.: Stereo disparity computation using Gabor filters.Biol.Cybernetics. 59,1988, 405-418.

[6] Krotkov E.P.: Active Computer Vision by Cooperative Focus and Stereo. Springer Verlag, New York,1989, 82ff.

[7] Franklin G. F.,Powell J. D.,Workman M. L.: Digital Control of Dynamic Systems. Addison Wesley, New York, 1990.

[8] Langley K., AthertonT.J., Wilson R.G., Larcombe M.H.E.: Vertical and Horizontal Disparities from Phase. Proc. 1st ECCV, 1990, 315-325.

[9] Fleet D.J., Jepson A.D., Jenkin, M.R.M.: Phased–Based Disparity Measurement. CVIGP Image Understanding, Vol.53, No.3, March,1991, 343-358.

[10] Olsen T.J.,Coombs D.J.: Real-Time Vergence Control for Binocular Robots. Internat. J. Computer Vision 7(1),1991, 76-89.

[11] Ballard D.H.,Brown C.: Principles of Animate Vision. CVGIP Image Understanding, Vol.56,1992, 3-21.

[12] Theimer W.M., Mallot H.A.: Binocular Vergence Control and Depth Reconstruction Using a Phase Method. Artificial Neural Networks, Vol.2,1992, 517-520.

[13] Michaelis M., Sommer G.: Basis functions for early vision. Techn. Report Nr. 9413, University Kiel,1994.

[14] Namuduri K.R., Mehrotra R., Ranganathan N.: Efficient Computation of Gabor filter based Multiresolution Responses. Pattern Recognition, Vol.27, No 7,1994, 925-938.

[15] Theimer W.M., Mallot H.A.: Phased–Based Binocular Vergence Control and Depth Reconstruction Using Active Vision. CVGIP Image Understanding, Vol.60, No.3, November,1994, 343-358.

[16] Westelius C.J., Knutsson H., Wiklund J., Westin C.F.: Phased-based Disparity Estimation. In Vision as Process, (Eds.) Crowley J.L., Christensen H.I., Springer Verlag, Heidelberg,1994.

[17] Sommer G.: Verhaltensbasierter Entwurf technischer visueller Systeme. Künstliche Intelligenz, No 3,1995, 42-45.

[18] Uhlin T., Nordlund P., Maki A., Eklundh J.O.: Towards an Active Visual Observer. International Journal of Pattern Recognition and Artificial CVAP, Stockholm, March 1995.