Stereo Matching with Implicit Detection of Occlusions

Ralph Trapp[1], Siegbert Drüe, and Georg Hartmann

University of Paderborn, Department of Electrical Engineering,
Pohlweg 47-49, D-33098 Paderborn, Germany
{trapp, druee, hartmann}@get.uni-paderborn.de

Abstract. In this paper we introduce a new stereo matching algorithm, in which the matching of occluded areas is suppressed by a self-organizing process. In the first step the images are filtered by a set of oriented Gabor filters. A complex-valued correlation-based similarity measurement, which is applied to the responses of the Gabor filters, is used in the second step to initialize a self-organizing process. In this self-organizing network, which is described by coupled, non-linear evolution equations, the continuity and the uniqueness constraints are established. Occlusions are detected implicitly without a computationally intensive bidirectional matching strategy. Due to the special similarity measurement, dense disparity maps can be calculated with subpixel accuracy. Unlike phase-difference methods the disparity range is not limited to the modulation wavelength of the quadrature-filter. Therefore, there is no need for a hierachical coarse-to-fine control strategy in our approach.

1 Introduction

Stereo vision is a passive method used to recover the depth information of a scene, which is lost during the projection of a point in the 3D-scene onto the 2D image plane. In stereo vision, in which two or more views of a scene are used, the depth information can be reconstructed from the different positions in the images to which a physical point in the 3D-scene is projected. The displacement of the corresponding positions in the image planes is called disparity. The central problem in stereo vision, known as the correspondence problem, is to find corresponding points or features in the images. This task can be an ambiguous one due to several similar structures or periodic elements in the images. Furthermore, there may be occluded regions in the scene, which can be seen only by one camera. In these regions there is no solution for the correspondence problem. Interocular differences such as perspective distortions, differences in illumination and camera noise make it even more difficult to solve the correspondence problem. Due to these aspects, the correspondence problem is a ill-posed problem according to Hadamard [5, 25]. The strategies used in solving the correspondence

1 This work is supported by a grant from the DFG Graduate Center „Parallele Rechnernetzwerke in der Produktionstechnik", ME 872/4-1.

problem can be divided into three major categories: area-based, feature-based and phase-based techniques. The area-based strategies use the intensities of the images to match them locally at each pixel [23]. Stereo techniques, which match features derived from intensity images rather than image intensities themselves, are called "feature-based" [12, 18, 21, 22, 26]. Whereas area-based methods produce dense disparity maps, feature-based strategies are only able to calculate disparities at locations where features occur. Because of the ill-posed character of the correspondence problem both techniques often yield multiple candidates for the match of a feature or an image area. Thus additional mechanisms are required to choose the correct match from among all the candidates. For this purpose techniques such as relaxation labeling [4, 18, 21, 27], regularization theory [3, 10, 11, 19, 25, 29] and dynamic programming [2, 24] are used by many researchers to impose global consistency constraints. Techniques in which disparity is expressed in terms of phase differences in the output of local band-pass filters applied to the images are called "phase-based" methods [7, 8, 15, 28]. The main advantage of phase-based techniques is that disparity estimations are obtained with subpixel accuracy without a subpixel feature localization. In contrast to area-based or feature-based strategies, an additional mechanism for the purpose of disambiguation is not necessary because the disparity range is limited to the modulation wavelength of the filter. To treat stereo images with a large disparity range, hierachical, strategies with multiple resolutions are necessary to obtain correct measurements. In such strategies the coarse disparity measurements are used as an initialization of the next finer level. If no special mechanisms are provided, all techniques of the three categories tend to produce wrong disparity estimates in occluded areas.

Without the occurrence of occluded regions in the images, stereo matching is a one to one mapping of the two images. In general, however, there may be several objects in the scene with different distances in relation to the cameras which cause discontinuity in disparity and occlusions near intensity edges defining the boundaries of different object surfaces. Constraints such as uniqueness, smoothness or ordering of the disparity, which are utilized to simplify the matching process are invalid assumptions in occluded regions. If occlusions are not specially treated in the matching process, they may be incorrectly matched with regions in the other image. Although occlusions are one of the essential reasons for wrong matches in stereo analysis, there are only a few approaches which treat them explicitly. One way to avoid correspondence errors in occluded areas is a bidirectional or dual matching process [14, 16, 20, 31]. In this approach matching is carried out from the left to the right and from the right to the left image in two separate, but identical processes. Occluded areas, which are indicated by the mismatch between the two disparity maps, are marked in so called occlusion maps and they are excluded from further calculations. This technique is based on the assumption that the match with features or pixel intensities in occluded areas is not as good as the match with the correct regions in the matching process, which is carried out in the other direction. Due to interocular differences, this need not be generally true. Another disadvantage of these techniques is the high computational complexity of the bidirectional matching process.

In our approach we use a complex-valued similarity measurement applied to the output of oriented Gabor filters [9]. Similar to phase-based methods, this measurement is very

robust with respect to interocular differences and provides dense data. Due to the correlation-based similarity measurement the disparity range is not limited, so coarse-to-fine control strategies are not required. We use the real part of this measurement to initialize a self-organizing process based on the pattern recognition equation introduced by Haken [13]. To each image point and to each possible disparity we assign a variable, which satisfies a non-linear evolution equation. The continuity constraint is established by a local cooperative coupling of the variables. All variables are involved in a competition so that after reaching the steady state of the dynamical process the variables with non-zero values represent the disparities at the image points. Contrary to the approach used by Reimann and Haken [27] the competition is arranged in a way that variables in occluded regions are prevented from winning the competition. Due to the special symmetry property of the similarity measurement no computationally intensive bidirectional match is required. After the correct disparities are determined by the self-organizing process, the imaginary part of the similarity measurement is used to improve the disparity estimation to subpixel accuracy.

2 Initial Processing

2.1 Gabor Filters

Since Sanger [28] proposed the use of the phase information in the output of local Gabor filters for binocular disparity measurements, many phase-based methods, which use quadrature-pairs of band-pass constant-phase filters, have been developed [7, 8, 15, 28]. The reason for the growing interest in these techniques are their numerous desirable properties. Disparity estimates are obtained with subpixel accuracy, dense disparity maps can be calculated, and no special treatment of the ambiguity of the correspondence problem is required. Furthermore, the measurements are robust with respect to smooth illumination differences between the two images because phase is amplitude invariant. In order to exploit some of these desirable properties, we decompose the images in the first step using a set of orientated Gabor filters.
Let $r(x)$ be the complex-valued result of the two-dimensional convolution of a Gabor filter $g(x)$ with an image $i(x)$ at the coordinate $x = [x_1, x_2]^T$.

$$r(x) = g(x) * i(x) \tag{1}$$

The Gabor filter $g(x)$ is tuned to an orientation ϕ and to a spatial frequency k_0.

$$g(x) = \frac{1}{2\pi ab} e^{-\frac{1}{2}x^T A x} e^{jk_0^T x} \tag{2}$$

where $j^2 = -1$ and

$$A = RPR^T = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} a^{-2} & 0 \\ 0 & b^{-2} \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}. \tag{3}$$

The spatial support of the filter and the bandwidth respectively depend on the parameter matrix P.

2.2 Similarity Measurement

There are some reasons why we do not use the phase-difference of the Gabor filters directly for disparity estimation: The basis for phase-difference methods consists of the Fourier shift theorem. But because of the local spatial support of the filters used in practice, the Fourier shift theorem does not strictly apply. Furthermore, the limited disparity range requires some form of coarse-to-fine control strategy, in which an initial guess is provided from coarser levels to bring the images into the disparity range of the next finer level. This common strategy may fail if the coarsest channel yields a poor estimate. In this case the process may converge to an incorrect disparity. To prevent errors due to phase-instabilities, which may occur when the filter output passes through the origin in the complex plane and to avoid coarse-to-fine control strategies, we use a correlation-based approach, which preserves the desirable properties of phase-based techniques.
With the convolution of the product of the left filter response $r_l(x)$ and a spatial shifted complex conjugate version of the right filter response $r_r(x)$ with a small real valued window $w(x)$, we obtain a local, complex-valued measurement $\rho_{lr}(x, d)$ of the similarity between the filtered images. This measurement is normalized to the local energy of the filter responses:

$$\rho_{lr}(x, d) = \frac{w(x) * r_l(x) r_r^*(x + d)}{\sqrt{w(x) * |r_l(x)|^2} \sqrt{w(x) * |r_r(x + d)|^2}} \tag{4}$$

where the disparity $d = [d_1, d_2]^T$ acts as a two-dimensional spatial displacement of the right filter response. Due to the epipolar constraint, the vertical component of the disparity is zero if the conventional parallel axis stereo geometry is used. If a non-parallel axis stereo geometry is used, the images can be easily transformed into parallel axes images by rectification [1]. In this case the convolution in (4) can be reduced to a one-dimensional convolution. The window $w(x)$ is chosen to be a one-dimensional gaussian with a support equal to the horizontal support of the Gabor filter.
The real and the imaginary parts of (4) can be expressed in terms of magnitude $|r(x)|$ and phase $\varphi(x)$.

$$\begin{aligned} \mathrm{Re}\{\rho_{lr}(x, d)\} &= \frac{w(x) * |r_l(x)||r_r(x + d)| \cos(\varphi_l(x) - \varphi_r(x + d))}{\sqrt{w(x) * |r_l(x)|^2} \sqrt{w(x) * |r_r(x + d)|^2}} \\ \mathrm{Im}\{\rho_{lr}(x, d)\} &= \frac{w(x) * |r_l(x)||r_r(x + d)| \sin(\varphi_l(x) - \varphi_r(x + d))}{\sqrt{w(x) * |r_l(x)|^2} \sqrt{w(x) * |r_r(x + d)|^2}} \end{aligned} \tag{5}$$

If the filter responses are locally similar, we expect a low phase difference. In this case there is a peak in the real part and we expect to find a zero-crossing in the imaginary

part due to the approximately linear phase of the Gabor filter output. The peaks in the real part of the similarity measurement act as candidate disparities d between the two images with pixel accuracy. Unlike some other approaches, which use the superposition of similarity measurements applied to several filter channels tuned to different frequencies as a kind of voting strategy [7], we use the real part of the measurement to initialize a self-organizing process to disambiguate the disparities. The zero-crossings of the imaginary part in the proximity of correct disparities are used to obtain subpixel accurate disparity estimations, which is described in chapter 4.
A further important property of the similarity measurement, which is central to the detection of occlusions, is the following identity between the measurement, which is done from the right image to the left image and the measurement which is done in the opposite direction.

$$\rho_{lr}(x, d) = \rho^*_{rl}(x + d, -d) \tag{6}$$

3 Correspondence by Self-organization

3.1 Related Work

One of the essential problems in stereo matching is that the correct correspondence may not be the one of the highest similarity. The reason for this is that the images may differ due to noise or distortions. Furthermore there may be points or areas in the image, which match equally well with several points or areas in the other images. These ambiguities, which are characteristic for ill-posed problems, can only be resolved by using natural constraints, which are general assumptions of the physical world. The most important constraints usually used in stereo vision are the uniqueness and the smoothness of the disparity map over the two-dimensional image plane first postulated by Marr and Poggio [21]. There are many approaches, which use these constraints to define energy functions, the global minimum of which is used to determine the correct correspondences. Strategies used to find the global minimum are, for instance, standard regularization theory [11, 25, 29] or stochastic relaxation [3, 10, 19].
To disambiguate the correspondence problem Reimann and Haken [27] use a kind of relaxation labeling algorithm. Similar to some other approaches of this category [4, 18] they assign a real-valued, time dependent variable $\xi(x, d, t)$ to each image coordinate x and to each possible disparity d. These variables, which are called binocular neurons, are involved in a dynamical process, which is described by coupled nonlinear evolution equations. The activity of these neurons is initialized by the output of an area-based matching technique. To satisfy the uniqueness constraint, the neurons are involved in a competition, which can be won only by one neuron at each image point. To get smooth disparity maps, each neuron is cooperatively connected to a small local area U representing the same disparity. The dynamics of the self-organizing process are described by the so-called coupled pattern recognition equations:

$$\dot{\xi}(x,d,t) = \Big\{\lambda(x,d) - (B+C)\sum_{d' \neq d} \xi^2(x,d',t) - C\xi^2(x,d,t) + \sum_{x' \in U} D(x')\xi(x',d,t)\Big\}\xi(x,d,t) \tag{7}$$

where the parameters B, C are positive constants. The first term $\lambda(x, d)$ represents an exponential growth of the amplitude $\xi(x, d, t)$ that depends on the area-based similarity between the images. The second term leads to a competition of all neurons at the same image point and the third term restricts the amplitude of $\xi(x, d, t)$. The fourth term represents the cooperative coupling, which is weighted by $D(x)$, of variables with the same disparity in a small local area in the image plane. It can be shown that only one of the variables at each image point reaches a non-zero stable fixed point and wins the competition in this way, while all of the other variables take the value zero. If we neglect the fourth term, we get the original pattern recognition equation introduced by Haken [13]. In this case the variable with the highest initial value always wins the competition. By adding the fourth term to the pattern recognition equation, variables, which have a strong cooperative area U, may win the competition even if they were initially smaller.

3.2 Initialization

In our approach, we use a modified version of the self-organizing process introduced by Reimann and Haken to disambiguate the disparity estimates resulting from the peaks in the real part of the similarity measurement defined in (4). For this purpose the variables of the dynamical process are initialized at the time $t = 0$ by a function f depending on the real part of the similarity measurement:

$$\xi(x,d,t=0) = f\Big(\sum_i c_i \mathrm{Re}\{\rho_{lr_i}(x,d)\}\Big) \tag{8}$$

where the subscript i denotes the similarity measurements, which are applied to several filter channels. These channels are tuned to different orientations (e.g. $-30°$, $0°$, $30°$) but to the same spatial frequency k_0. The channels i can be weighted individually by the parameters c_i. The function f is used to map the values of the real part of the similarity measurement to a positive range by truncating the negative values. In the following sections the variables are involved in a special self-organizing process, in which natural constraints are exploited to solve the correspondence problem.

3.3 Treatment of Occluded Regions

Since there is no special treatment of occlusions in the approach of Reimann and Haken, an initialization of a variable in an occluded area with a non-zero value always results in a wrong disparity estimation, because there is no solution of the correspon-

dence problem in this area. In order to prevent those correspondence errors, we define a new competition in a self-organizing process, which suppresses variables in occluded image regions.

Occlusions can be detected by a similarity measurement, which is carried out from left to right and vice versa. Figure 1 shows a random dot stereogram, in which a square area marked by the black frame is inserted in the images with a relative shift of ten pixels. The image point marked by the white point in the left image does not occur in the right image. The similarity measurement, which is carried out from left to right at this point shows a peak in the real part, which corresponds to the position marked by the black point in the right image. Due to its high similarity measurement this wrong disparity would probably win the competition. As a result of (6) the real part of the similarity measure carried out from right to left must show the same peak at the disparity with the opposite sign at the corresponding position in the right image. But in this measurement there is additionally a higher peak, which corresponds to the match with the correct image area in the left image marked by the black point. If the values of this measurement are also included in the competition, the disparity with the highest peak in this measurement is likely to win.

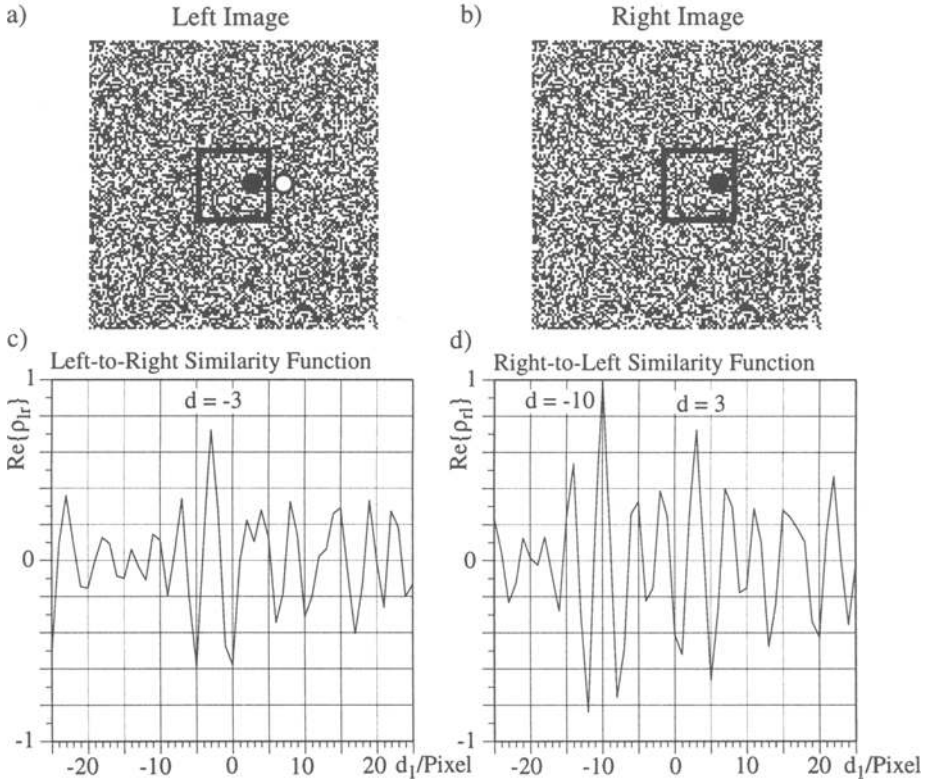

Fig. 1. a) and b) Random dot stereogram. c) Real part of the left-to-right similarity function of an occluded area marked by the white point in the left image. d) Real part of the right-to-left similarity function at the position marked by the black point in the right image, which corresponds to the maximum in c)

To prevent the match of occluded areas with other image points, we include both directions of the similarity measurement in the competition. By exploiting equation (6), which is also valid for the variables $\xi(x, d, t)$, the new competition term is given by:

$$\sum_{d' \neq d} \xi^2(x, d', t) + \xi^2(x + d - d', d', t) \tag{9}$$

Contrary to other bidirectional matching techniques [14, 16, 20, 31] the occlusion detection in our approach is done implicitly by a matching process, which is carried out in only one direction. Therefore, an explicit calculation of a computationally intensive bidirectional match is not necessary. Without loss of generality we refer in the following part only to the match, which is done from left to right.

3.4 Treatment of Matches between Occlusions

If a cooperative term as in (7) is used with a competitive term as shown in (9) a match of occluded image regions with areas seen from both views can be prevented even if the wrong match is locally higher than the correct match done in the opposite direction. Unfortunately, an object may produce occluded areas in both images, which occur on different sides of the object. So, if the established disparity range is large enough and a match between this occluded areas can be calculated, it is probable that a variable will win the competition, which represents a correspondence between the occluded areas. To treat this type of wrong correspondence, we consider a cyclopean camera, which was introduced by Julesz [16]. From the point of view of a cyclopean camera, the reconstructed 3D-coordinates of the correspondence between these occluded areas are found either in front of or behind the object that produces these occlusions. The cyclopean image coordinates x_c of a variable at the coordinates x in the left image and with disparity d are given by:

$$x_c = x + \frac{d}{2} \tag{10}$$

In this cyclopean image the variables representing the object and the correspondence between the occlusions have the same image coordinates. In order to suppress correspondences between occlusions, we extent the self-organizing process once again by involving all variables in the same competition, which show the same cyclopean image coordinates. This strategy can be implemented by using the following competition term:

$$\sum_{d' \neq d} \xi^2(x, d', t) + \xi^2(x + d - d', d', t) + \xi^2\left(x + \frac{d}{2} - \frac{d'}{2}, d', t\right) \tag{11}$$

Because occluded areas are unlikely to match better in large image areas than correct correspondences, matches between occlusions are suppressed by this technique. The dynamical behavior of the resulting self-organizing process is given by:

$$\begin{aligned}\dot{\xi}(x,d,t) = \Big\{ & A - C\xi^2(x,d,t) \\ & - \frac{B}{3N}\sum_{d' \neq d} \xi^2(x,d',t) + \xi^2(x+d-d',d',t) + \xi^2\Big(x+\frac{d}{2}-\frac{d'}{2},d',t\Big) \\ & + \frac{D}{M}\sum_{x' \in U} \xi(x',d,t)\Big\}\xi(x,d,t) \\ = {} & F\{x,d,t\}\end{aligned} \tag{12}$$

where the parameters A, B, C, D are positive constants. The terms that lead to the competition and to the cooperative coupling are normalized to the amount of involved variables N, M. For reasons of simplicity we choose the coupling constant D to be equal to all variables in the cooperative area U and the exponential growing term to be independent of the similarity measurement. In practice this self-organizing process can be easily implemented as an iterative algorithm if the differential equations are integrated numerically.

3.5 Properties of the Self-organizing Process

The parameters in equation (12) have to be chosen suitably in order to establish a competition and to insure the stability of the fixed points of the system. The fixed points $\xi_r(x,d)$ of (12) satisfy the condition $\dot{\xi}_r(x,d,t) = 0$. There are three possible solutions for the fixed points:

$$\begin{aligned}\xi_{r_1}(x,d) &= 0 \\ \xi_{r_{2/3}}(x,d) &= \pm\sqrt{\frac{\Gamma(x,d)}{C}}\end{aligned} \tag{13}$$

where $\Gamma(x,d)$ is given by:

$$\begin{aligned}\Gamma(x,d) = \Big(& A + \frac{D}{M}\sum_{x' \in U} \xi_r(x',d) \\ & - \frac{B}{3N}\sum_{d' \neq d} \xi_r^2(x,d') + \xi_r^2(x+d-d',d') + \xi_r^2\Big(x+\frac{d}{2}-\frac{d'}{2},d'\Big)\end{aligned} \tag{14}$$

By initializing the process with positive values, it can be shown (see appendix A) that the negative fixed points in (13) are never reached, thus we only consider the fixed points $\xi_{r_1}(x,d)$ and the positive fixed points $\xi_{r_2}(x,d)$.
Under certain conditions a competition can take place, in which variables with a strong cooperative support and a high initial value are likely to win (see appendix B). The winner of the competition is the variable that takes the stable fixed point solution $\xi_{r_2}(x,d)$. This fixed point should be reached only by one variable whereas all other variables, which are involved in the competition, should decrease to the fixed point solution $\xi_{r_1}(x,d)$. Furthermore, the parameters have to be chosen in a way that the sta-

bility of the fixed points is guaranteed under certain circumstances (see appendix A). The influence of the variables in the cooperative area on the self-organizing process is steered by the parameter D.

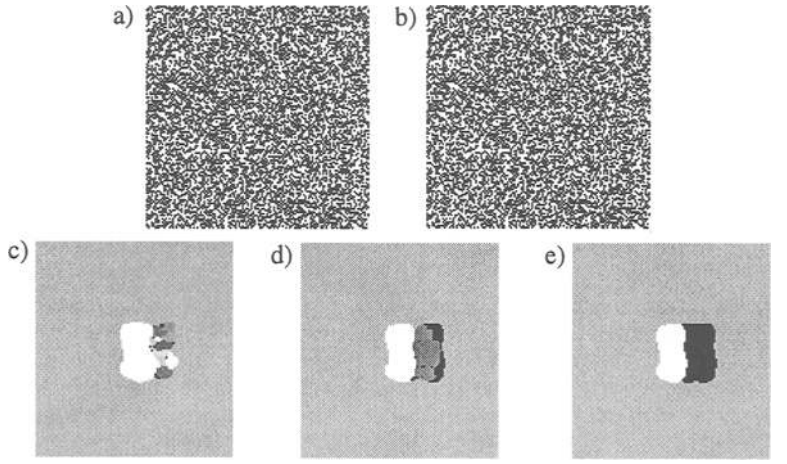

Fig. 2. a) and b) 50% Random dot stereogram. c) Disparity map without occlusion treatment. d) Disparity map generated by a self-organizing process with an extended competition defined by equation (9). e) Disparity map generated by using equation (12)

Figure 2 shows a random dot stereogram in which a rectangular area is inserted in the images with a relative displacement, which is equal to the width of the inserted rectangle. The results of the self-organizing process with different competition terms but equal parameters are shown by the disparity maps below the stereogram. The gray values in the disparity maps represent the disparities of the variables, which have reached the fixed point solution $\xi_{r_2}(x, d)$. If there is no variable at an image point that has reached the solution $\xi_{r_2}(x, d)$ the disparity remains undefined. These areas are marked black in the disparity maps. The disparity range was chosen to be twice as large as the width of the displaced rectangle. The left disparity map shown in figure 2c is calculated by a self-organizing process defined in (7). Because there is no special treatment of occlusions, the disparities on the right of the rectangle, represent false, accidental matches of the occluded area, because there are no correct correspondence partners in the right image.

The disparity map shown in figure 2d is generated by using an extended competition term defined in (9). In the occluded area only those variables have reached the fixed point $\xi_{r_2}(x, d)$, which represent matches with the occluded region in the right image. Because the disparities of these wrong matches are negative in this example, they are darker than the disparity of the rectangle. In the black areas none of the variables, which represent a match between the occluded areas in both images, have been initialized to a non-zero value by the function defined in (8). The disparity map in figure 2e is obtained by using the new self-organizing process introduced in (12). Due to the competition term defined in (11), all variables in the occluded area have been suppressed, and no disparity value has been calculated.

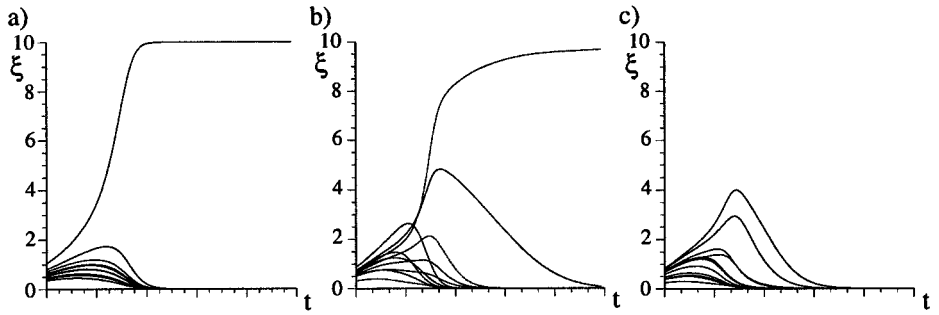

Fig. 3. Time dependent behavior of variable values at different image points. a) Image area without ambiguous matches. The variable with the highest initial value wins the competition. b) Image area with ambiguous matches. Due to the local cooperative coupling, a variable wins the competition, which was not the initially highest one. c) Competition in an occluded image area. No variable reaches the non-zero fixed point

The competition of variables lying at the same cyclopean image coordinates, which is established by (11), is related to the disparity gradient limit first proposed by Burt and Julesz [6]. In several approaches an approximation of the disparity gradient limit is used to disambiguate the correspondence problem (see for example [26]). Usually the disparity gradient is approximated by:

$$D_{pq} = \left| \frac{d_p - d_q}{x_p - x_q} \right| \tag{15}$$

where $x_{p/q}$ are the cyclopean image coordinates of the points p, q and $d_{p/q}$ the corresponding disparities. A match is only permitted if the disparity gradient is lower than a certain limit. For $D_{ij} < 2$ the disparity gradient limit is equal to the ordering constraint, which is often used in approaches using dynamic programming techniques (see for example [2]). In the steady state of our self-organizing process described in (12) there is at most one variable at the same cyclopean image point that reaches a non-zero fixed point. Therefore, matches are only prevented in our approach if their disparity gradient is infinity. This restriction of the disparity is much weaker than the ordering constraint or the disparity gradient limits used in practice.

4 Disparity Estimation with Subpixel Accuracy

As mentioned in chapter 2.2 the imaginary part of the similarity measurement introduced in (4) can be exploited to obtain a disparity estimation with subpixel accuracy. Due to the approximate linear phase of the Gabor filter, we expect to find a zero-crossing in the imaginary part of the similarity measurement close to a peak in the real part. After the self-organizing process has reached the steady state we interpolate the imaginary part at the position of variables with non-zero values, because they are likely to correspond to a peak in the real part of the similarity measurement. Then we use the position of the zero-crossings in the imaginary part as the new subpixel accurate dis-

parity estimations. By using this technique we are able to obtain subpixel accurate disparity estimations without blurring of the discontinuities in the disparity map. The best results are obtained when the imaginary part of the similarity measurement is used, which gets its input from the filter orientation, in which the highest magnitude of the filter response occurs. Figure 4 shows a stereogram of white noise, in which one image is a copy of the other image, which is stretched horizontally by 10%.

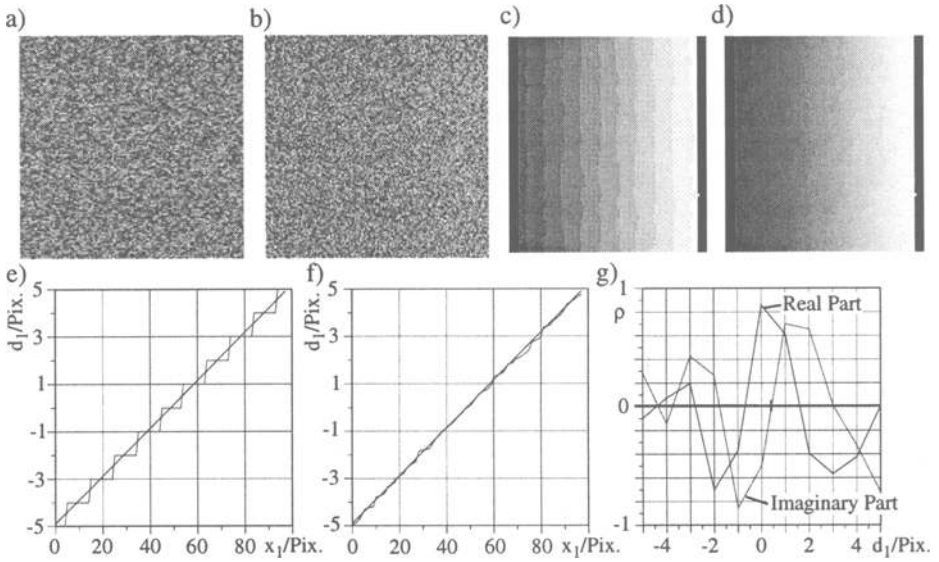

Fig. 4. a) + b) Stereogram generated with white noise. c) + e) Disparity map and disparity function with pixel accuracy. d) + f) Disparity map and disparity function with subpixel accuracy. g) Real and imaginary part of the similarity measurement

If the disparity is calculated by interpolating the zero-crossings in the imaginary part of the similarity measurement as shown in figure 4g, the root mean square of the quantization errors can be reduced in this example by 60%.

5 Experimental Results

The current algorithm has been tested on a number of real and artificial image pairs. In this section some of the results are presented. The parameters were chosen identically for all examples to $A = 0.4$, $B = 1.0$, $C = 0.01$, and $D = 0.61$. The cooperative area U was chosen to be a square area of 5x5 pixel. To reduce the computational expenditure, the horizontal disparity range was restricted to $|d_1| \leq 15$ pixel. The presented natural image pairs had been rectified, thus no vertical disparities had to be calculated. The Gabor filters were tuned to a modulation frequency of $k_0 = \pi/2$ and had a bandwidth of 0.6 octaves. The images were filtered by three channels tuned to the orientations $(-30°, 0°, 30°)$. The values of the superposed similarity measurements of the filter channels were truncated by their negative results and normalized to the number of

channels by the function f to initialize the variables of the self-organizing process. Detected occlusions are marked black in the disparity maps. The disparity maps are registered in the right image coordinates.

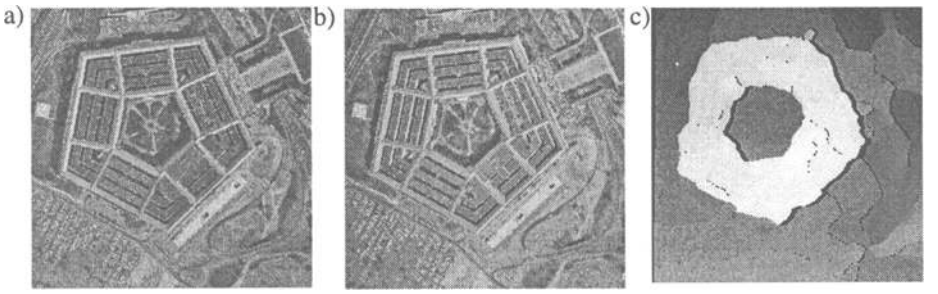

Fig. 5. a) Left and b) right image of the "Pentagon" stereo pair. c) Disparity map

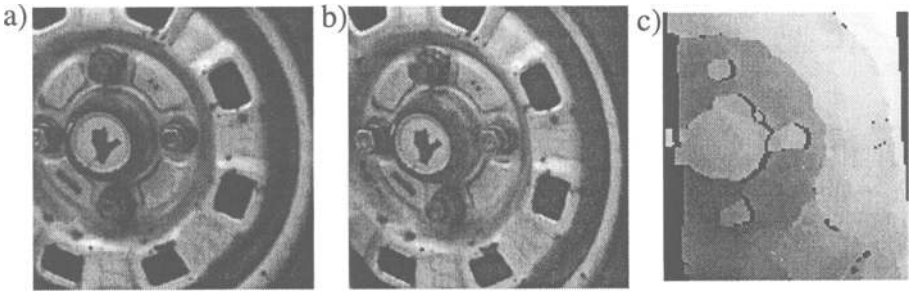

Fig. 6. Rectified a) left and b) right image of an old car tire. d) Disparity map

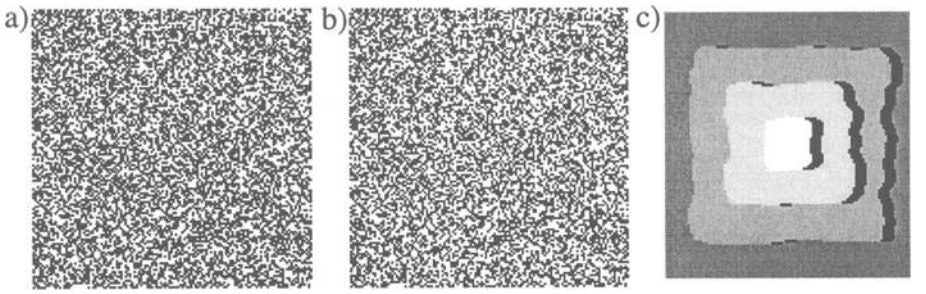

Fig. 7. a) Left and b) right image of a 50% random dot stereogram with four square areas of different disparity. c) Disparity map

6 Conclusion

In this paper we have outlined a new method for the estimation of disparities in stereo image pairs. In our approach we have combined some desirable properties of existing techniques such as the subpixel accuracy of phase-based techniques and the robustness and the high parallelization potential of cooperative approaches. Furthermore, we have

implemented a new method of detecting occlusions in stereo images, which are one of the most important error sources in stereo matching. Due to a complex-valued, correlation-based technique, which is applied to the output of oriented Gabor filters, the algorithm is robust to smooth illumination differences between the images and provides dense, subpixel accurate disparity estimations. The continuity constraint and an extended version of the uniqueness constraint are applied in a self-organizing process. Because of this new technique occlusions are detected implicitly and no disparity estimates are calculated in this areas. As shown in some examples the algorithm produces reliable results in artificial and natural images, even if there are large occluded areas or disparity gradients in them.
Future work will include a fast, parallel implementation of this algorithm for the purpose of distance measurement and obstacle avoidance in order to navigate an autonomous robot manipulator.

Appendix A: Stability of the Fixed Points

To guarantee the stability of the fixed points, we have to consider the functional matrix of the linearized system. The non-zero elements of the matrix are:

$$\begin{aligned}
\frac{\partial}{\partial\xi(x,d,t)}F(x,d,t)|_{\xi_r} &= \Gamma(x,d) - 3C\xi_r^2(x,d) \\
\frac{\partial}{\partial\xi(x,d',t)}F(x,d,t)|_{\xi_r} &= -2B\xi_r(x,d')\xi_r(x,d)/3N \quad \text{for } d' \neq d \\
\frac{\partial}{\partial\xi(x+d-d',d',t)}F(x,d,t)|_{\xi_r} &= -2B\xi_r(x+d-d',d')\xi_r(x,d)/3N \quad \text{for } d' \neq d \\
\frac{\partial}{\partial\xi\left(x+\frac{d}{2}-\frac{d'}{2},d',t\right)}F(x,d,t)|_{\xi_r} &= -2B\xi_r\left(x+\frac{d}{2}-\frac{d'}{2},d'\right)\xi_r(x,d)/3N \text{ for } d' \neq d \\
\frac{\partial}{\partial\xi(x',d,t)}F(x,d,t)|_{\xi_r} &= \frac{D}{M}\xi_r(x,d) \quad \text{for } x' \in U
\end{aligned} \tag{16}$$

We now consider the stability of the fixed points under several different conditions. In the first case, we assume all variables of the competition to take the fixed point solution $\xi_{r_1}(x,d) = 0$. The variables of the cooperative area may take any value. Under this conditions the functional matrix has a diagonal sub-block of the form:

$$\frac{\partial}{\partial\xi(x,d,t)}F(x,d,t)|_{\xi_{r_1}} = A + \frac{D}{M}\sum_{x' \in U}\xi_r(x',d) \tag{17}$$

Due to the fact that A and D are positive constants and all variables are positive, this kind of fixed points is not stable, because the eigenvalues of this sub-block are positive. Thus not all of the variables, which are involved in the same competition process, can take the fixed point solution $\xi_{r_1}(x,d) = 0$ if they were initialized by a non-zero value.

If one of the variables takes the solution $\xi_{r_1}(x, d)$, whereas all other variables in the same competition process may take any non-negative value, the fundamental matrix has a diagonal sub-block, which is given by:

$$\frac{\partial}{\partial \xi(x, d, t)} F(x, d, t)|_{\xi_{r_1}} = \Gamma(x, d) \tag{18}$$

This fixed point is only stable if $\Gamma(x, d)$ is negative, otherwise it is unstable. Therefore, the variables are unable to take negative values if they are initialized by positive or zero values. Thus, the negative fixed point solution of (13) is never reached. We want to choose the parameters in a way that this kind of fixed points is stable if one variable in the competition process reaches the fixed point solution $\xi_{r_2}(x, d)$, which is given by (13). This leads to the inequality:

$$A - \frac{B}{3N} \xi_{r_2}^2(x, d) < 0 \tag{19}$$

The value of $\xi_{r_2}(x, d)$ can be approximated by the worst case, in which no variable in the cooperative area takes a non-zero value. The non-zero fixed point is then given by $\xi_{r_2}(x, d) = \sqrt{A/C}$. Under these conditions the parameters B, C, and N must fulfill the following inequality:

$$C < \frac{B}{3N} \tag{20}$$

The fixed point $\xi_{r_2}(x, d)$ should be stable if all other variables in the competition process have reached the fixed point solution $\xi_{r_1}(x, d)$. In this case the non-vanishing elements of this sub-block of the functional matrix are given by:

$$\begin{aligned} \frac{\partial}{\partial \xi(x, d, t)} F(x, d, t)|_{\xi_{r_2}} &= -2\Gamma(x, d) \\ \frac{\partial}{\partial \xi(x', d, t)} F(x, d, t)|_{\xi_{r_2}} &= \frac{D}{M} \xi_{r_2}(x, d) \quad \text{for } x' \in U \end{aligned} \tag{21}$$

From the negative diagonal elements of this sub-block it follows using the theorem of Gerschgorin, that the eigenvalues of this sub-block are negative when the sum of the non-diagonal elements in each row is smaller than every diagonal element (see for example [30]). This is generally true if at least half of the variables in the cooperative area have reached also the fixed point solution $\xi_{r_2}(x, d)$.

Appendix B: Further Parameter Restrictions

To enable a competition, in which the variable with the greatest support of the cooperative area and with the highest value grows stronger than the other involved variables, the parameters need a further restriction.

Let $\xi(x_1, d_1, t)$ be the variable with the highest value in the competition and with the highest cooperative support $u(x_1, d_1, t) = \sum_{x' \in U_1} \xi(x', d_1, t)$,

$$\dot{\xi}(x_1, d_1, t) > \dot{\xi}(x, d, t) \qquad \forall d \neq d_1 \text{ and } x \in \left\{ x_1, x_1 - d_1 - d, x_1 - \frac{d_1}{2} - \frac{d}{2} \right\}, \quad (22)$$

then follows from (12):

$$\begin{aligned} &\xi(x_1, d_1, t)\left(A - \frac{B}{3N}s + \frac{B}{3N}\xi^2(x_1, d_1, t) - C\xi^2(x_1, d_1, t) + \frac{D}{M}u(x_1, d_1, t)\right) > \\ &\xi(x, d, t)\left(A - \frac{B}{3N}s + \frac{B}{3N}\xi^2(x, d, t) - C\xi^2(x, y, t) + \frac{D}{M}u(x, d, t)\right) \end{aligned} \quad (23)$$

where s is the sum of the squares of all variable values involved in the competition.

$$s = \sum_{d'} \xi^2(x_1, d', t) + \xi^2(x_1 + d_1 - d', d', t) + \xi^2\left(x_1 + \frac{d_1}{2} - \frac{d'}{2}, d', t\right) \quad (24)$$

Because the sum s is equal to all involved variables and the cooperative support of $\xi(x_1, d_1, t)$ is higher than the support of the other involved variables, the inequality (23) is fulfilled, if the following inequality is valid:

$$\left(\frac{B}{3N} - C\right)\xi^2(x_1, d_1, t) > \left(\frac{B}{3N} - C\right)\xi^2(x, d, t). \quad (25)$$

Under the assumption made above this results in the parameter restriction:

$$\frac{B}{3N} > C. \quad (26)$$

This inequality is always fulfilled if the stability of the fixed point $\xi_{r_1}(x, d)$ is guaranteed for the worst case discussed in appendix A (see equation (20)).

References

1. Ayache, N.; Hansen, C.: Rectification of Images for Binocular and Trinocular Stereovision. In Proc. of the 9th Intern. Conference on Pattern Recognition, pp. 11-16, Oktober 1988.
2. Baker, H. H.; Binford, T. O.: Depth from Edge and Intensity Based Stereo. In Proceedings of the 7th International Joint Conference on Artifical Intelligence, pp. 631-636, August 1981.
3. Barnard, S. T.: Stochastic stereo matching over scale. In International Journal of Computer Vision, vol. 2, pp. 17-32, 1989.
4. Barnard, S. T.; Thompson, W. B.: Disparity Analysis of Images. In Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-2, no. 4, pp. 333-340, July 1980.
5. Bertero, M.; Poggio, T. A.; Torre, V.: Ill-Posed Problems in Early Vision. In Proceedings of the IEEE, vol. 76, no. 8, pp. 869-889, August 1988.
6. Burt, P.; Julesz, B.: A Disparity Gradient Limit for Binocular Fusion. In Science, vol. 208, pp. 615-617, 1980.
7. Fleet, D. J.: Disparity from Local Weighted Phase-Correlation. In IEEE International Conference on Systems, Man and Cybernetics, pp. 48-56, San Antonio, October 1994.
8. Fleet, D. J.; Jepson, A. D.; Jenkin, M. R. M.: Phase-Based Disparity Measurement. In CVGIP: Image Understanding, vol. 53, no. 2, pp. 198-210, March 1991.

9. Gabor, D.: Theory of communication. In J. IEE, vol. 93, pp. 429-457, London, 1946.
10. Geman, S.; Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pp. 721-741, November 1984.
11. Gennert, M. A.: Brightness-based stereo matching. In International Conference on Computer Vision, pp. 139-143, 1988.
12. Grimson, W. E. L.: A computer implementation of a theory of human stereo vision. In Philosophical Transactions of the Royal Society of London, vol. B 292, pp. 217-253, 1981.
13. Haken, H.: Synergetic Computers and Cognition. Springer-Verlag, Berlin, 1991.
14. Hoff, W.; Ahuja, N.: Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 2, February 1989.
15. Jenkin, M. R. M.; Jepson, A. D.: Recovering Local Surface Structure through Local Phase Difference Measurements. In Computer Vision, Graphics and Image Processing: Image Understanding, vol. 59, no. 1, pp. 72-93, January, 1994.
16. Jones, D. G.; Malik, J.: A computational framework for determining stereo correspondence from a set of linear spatial filters. In Proceedings of the 2nd European Conference on Computer Vision, Springer Verlag, Berlin, pp. 395-410, 1992
17. Julesz, B.: Foundations of cyclopean perception. University of Chicago Press, 1971.
18. Kim, Y. C.; Aggarwal, J. K.: Positioning Three-Dimensional Objects Using Stereo Images. In IEEE Journal of Robotics and Automation, vol. RA-3, no. 4, pp. 361-373, August 1987.
19. Konrad, J.; Dubois, E.: Multigrid Bayesian Estimation of Image Motion Fields Using Stochastic Relaxation. In Proceedings of the 2nd International Conference on Computer Vision, pp. 354-362, 1988.
20. Luo, A.; Burkhardt, H.: An Intensity-Based Cooperative Bidirectional Stereo Matching with Simultaneous Detektion of Discontinuities and Occlusions. In International Journal of Computer Vision, vol. 15, pp. 171-188, 1995.
21. Marr, D.; Poggio, T.: A computational theory of human stereo vision. In Proceedings of Royal Society of London, vol. B 204, pp. 301-328, 1979.
22. Medioni, G.; Nevatia, R.: Segment-Based Stereo Matching. In Computer Vision, Graphics, and Image Processing, vol. 31, pp. 2-18, 1985.
23. Moravec, H. P.: Towards automatic visual obstacle avoidance. In Proceedings of the 5th International Joint Conference on Artifical Intelligence, p. 584, 1977.
24. Otha, Y.; Kanade T.: Stereo by intra- and inter-scanline search. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-7, no. 2, pp. 139-154, March 1985.
25. Poggio, T. A.; Torre, V.; Koch, C.: Computational Vision and Regularization Theory. In Nature, vol. 317, pp. 314-319, 1985.
26. Pollard, S. B.; Mayhew, J. E. W.; Frisby, J. P.: PMF: A stereo correspondence algorithm using a disparity gradient constraint. In Perception, 14, pp. 449-470, 1985.
27. Reimann, D.; Haken, H.: Stereovision by Self-Organisation. In Biological Cybernetics, vol. 71, pp. 17-26, 1994.
28. Sanger, T. D.: Stereo Disparity Computation Using Gabor Filters. In Biological Cybernetics, vol. 59, pp. 405-418, 1988.
29. Terzopoulos, D.: Regularization of Inverse Visual Problems Involving Discontinuities. In IEEE Trans. Pattern Anal. and Machine Intel., vol. PAMI-8, no. 4, pp. 413-424, July 1986.
30. Varga, R. S.: Matrix Iterative Analyse. G. Forsythe (ed.), Prentice-Hall Series in Automatic Computation, New Jersey, 1962.
31. Weng, J.; Ahuja, N.; Huang, T. S.: Two-view matching. In Proceedings of the 2nd International Conference on Computer Vision, pp. 64-73, 1988.