Biol. Cybern. 78, 329–336 (1998)

Biological Cybernetics

How to "hear" visual disparities: real-time stereoscopic spatial depth analysis using temporal resonance

Bernd Porr, Alex Cozzi, Florentin Wörgötter

Institut für Physiologie, Ruhr-Universität Bochum, D-44780 Bochum, Germany

Received: 9 April 1997 / Accepted in revised form: 15 January 1998

Abstract. In a stereoscopic system, both eyes or cameras have a slightly different view. As a consequence, small variations between the projected images exist ('disparities') which are *spatially* evaluated in order to retrieve depth information (Sanger 1988; Fleet et al. 1991). A strong similarity exists between the analysis of visual disparities and the determination of the azimuth of a sound source (Wagner and Frost 1993). The direction of the sound is thereby determined from the *temporal* delay between the left and right ear signals (Konishi and Sullivan 1986). Similarly, here we transpose the spatially defined problem of disparity analysis into the temporal domain and utilize two resonators implemented in the form of causal (electronic) filters to determine the disparity as local temporal phase differences between the left and right filter responses. This approach permits real-time analysis and can be solved analytically for a step function contrast change, which is an important case in all real-world applications. The proposed theoretical framework for spatial depth retrieval directly utilizes a temporal algorithm borrowed from auditory signal analysis. Thus, the suggested similarity between the visual and the auditory system in the brain (Wagner and Frost 1993) finds its analogy here at the algorithmical level. We will compare the results from the temporal resonance algorithm with those obtained from several other techniques like cross-correlation or spatial phase-based disparity estimation showing that the novel algorithm achieves performances similar to the 'classical' approaches using much lower computational resources.

Correspondence to: F. Wörgötter
(e-mail: worgott@neurop.ruhr-uni-bochum.de,
Fax: +49-234-709-4192)

Supplementary material: A colored version of figure 6 has been deposited in electronic form and can be obtained from http://link.Springer.de/link/service/journals/00422/tocs.htm

1 Introduction

The field of biological cybernetics and neural modeling has undergone several transitions over the last decades. Classical 'cybernetical' approaches which dominated before 1970 (often involving linear systems theory) were soon followed by neuronal network models with different degrees of biological realism. The domain of artificial neural networks (ANN) began to exert its massive influence in the last 10 years or so. The strongest driving force behind ANN research was probably the attempt to transfer ideas taken from biology to a more technological domain. Thus, this aspect of neuronal modeling (in its widest sense) was especially attractive to engineers and other application-oriented researchers. As a consequence, at about this time neural modeling 'became useful' also outside the field of brain science. The transfer of biological ideas to technology, however, is not necessarily restricted to ANNs, and this may be a sensible consideration given the intrinsic disadvantages of ANNs (e.g., slow relaxation behavior). Instead, sometimes it is possible to design an application-oriented algorithm in a rather direct way from a biologically inspired model.

Thus, the goal of this article is twofold: we will try to show that (1) an algorithm stolen from the auditory system can be applied to a visual problem, and (2) that it is possible to transfer this algorithm directly to a chain of electronic filters which operate in real-time. To this end, we will concentrate on the problem of stereo-image analysis.

In any vision-based system the 3-dimensional world is projected onto 2-dimensional receptor surfaces. These could be the two retinas of a binocularly viewing animal or the cameras of an artificial system. During that process, depth information is lost but can be recovered from the disparities between matching image parts. In technical systems, vertical disparities are often neglected by assuming a strictly frontoparallel camera geometry. In this case, it is sufficient to analyze corresponding cross-sections of both images line by line because the

epipolar lines are now horizontal. Thus, stereoscopic depth estimation is reduced to a 1-dimensional spatial problem, and common methods use acausal spatial filters to retrieve the disparity as a convolution result (Sanger 1988; Fleet et al. 1991). The inherently present restriction to one dimension, however, makes it also possible to interpret each line from the left and right image as a temporal signal $x(t)$, which could, for example, be imagined as scan-line from a CCD camera arriving pixel after pixel. With the help of this interpretation, a causal filter approach can now be defined such that the disparity is detected continuously with the incoming data.

2 Causal filtering of the stereo-images

2.1 General description

The system we present is very simple: It takes the luminance signal of the image scan-lines from the left and the right image and pipes it through a left and a right band-pass filter (a resonator). This way two signals are generated which are quasi-oscillatory at the resonance frequency. The (local) phase difference between these two oscillations is directly equivalent to the disparity. Thus, subsequently our system measures this phase difference by two more simple electronic operations as shown in Fig. 1 and explained in the next section.

2.2 Equations for the generic case

We assume a frontoparallel camera arrangement, which leads to horizontal epipolar lines. Disparity changes can only be detected when they concur with a luminance change. For digitized camera data, the smallest luminance change is a 1-bit step function. In addition, stronger step-like luminance changes in general occur rather often in images, for example, at the edges of a protruding object. Thus, step functions are a very generic case for which we will solve the 'Ansatz' analytically. Let $x_\mathrm{l}(t)$, $x_\mathrm{r}(t)$ be the two corresponding pixel lines of a stereo image pair in which a single contrast step exists at different disparities (viz. different times t_l and t_r). To obtain the disparity between the images, each signal is used to excite a resonator with characteristic frequency f_0. We assume that the contrast step in the left image occurs earlier than that of the right image ($t_\mathrm{l} < t_\mathrm{r}$), and thus that the resulting resonance starts earlier for x_l than for x_r. This temporal phase difference is directly equivalent to the spatial disparity between the images and can be obtained from an operator which compares the phases.

The two step functions $x_\mathrm{l}(t) \leftrightarrow X_\mathrm{l}(s)$ and $x_\mathrm{r}(t) \leftrightarrow X_\mathrm{r}(s)$ are defined in the Laplace domain by (Fig. 1):

$$X_\mathrm{l}(s) := \tfrac{1}{s}\,\mathrm{e}^{-t_\mathrm{l}s} \text{ and } X_\mathrm{r}(s) := \tfrac{1}{s}\,\mathrm{e}^{-t_\mathrm{r}s} \tag{1}$$

and the transfer function of the resonator is given as:

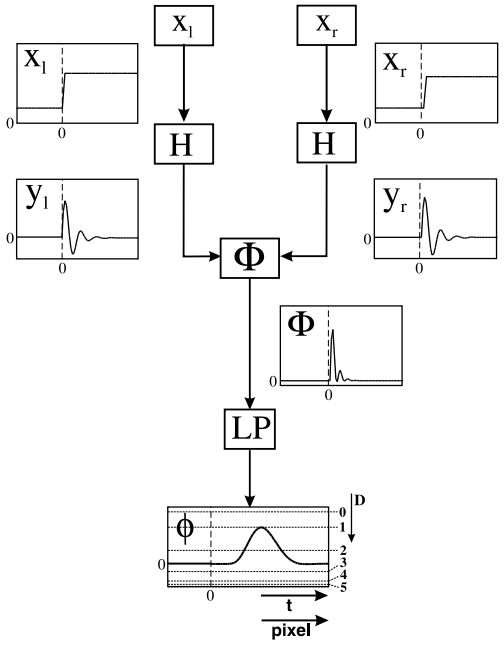

Fig. 1. Block diagram of the computational process and results of a disparity estimation from the two input step functions x_r and x_l. The initial disparity was 1 pixel, f_0 was 0.1 pixel^{-1}, and for graphical reasons we have set Q to 2.0 such that two full oscillation cycles are shown. $y_\mathrm{r}, y_\mathrm{l}$ reflect the resonator responses (Eq. 5), Φ is the signal from the phase comparison (Eq. 6) and ϕ shows the disparity output after low-pass filtering (Eq. 8). The scaling of ϕ reflects the damped cosine characteristic of Eq. (9). In this technical implementation, the constant delay until read-out of the disparity was 10.0 pixels. For a CCD camera-based system, we can assume a pixel input rate of >10 MHz. Thus, the final output of this disparity processing system would be available after a total delay of only 1.00 μs. The measured disparity after that delay, i.e., at the peak of ϕ, is 0.97 pixels

$$H(s) = \frac{s}{(s - s_\infty)(s - s_\infty^*)} \tag{2}$$

where s_∞ is a filter pole and specifies the filter characteristic defined by f_0 and the filter quality Q, which determines the damping; the asterisk denotes the complex conjugate.

$$\mathrm{Re}(s_\infty) = -2\pi f_0/2Q;\;\; \mathrm{Im}(s_\infty) = \sqrt{(2\pi f_0)^2 - (\mathrm{Re}(s_\infty))^2} \tag{3}$$

Convolution of signal and filter yields for the right image:

$$Y_\mathrm{r}(s) = X_\mathrm{r}(s)H(s) = \frac{s}{(s - s_\infty)(s - s_\infty^*)}\frac{1}{s}\mathrm{e}^{-t_\mathrm{r}s} \tag{4}$$

A similar convolution is performed for the left image.

We define $a := (s_\infty - s_\infty^*)^{-1}$, then the inverse Laplace transformation of $Y_\mathrm{r}(s)$ yields:

$$y_r(t) = \begin{cases} a\mathrm{e}^{s_\infty(t-t_\mathrm{r})} + a^*\mathrm{e}^{s_\infty^*(t-t_\mathrm{r})} & \text{if } t \geq t_\mathrm{r} \\ 0 & \text{if } t < t_\mathrm{r} \end{cases} \tag{5}$$

The temporal resonator signal $y(t)$ reflects a damped sine-wave with frequency f_0 (Fig. 1, y_l, y_r). The number

of full cycles until the signal fades is roughly equivalent to the value of Q. Note that any DC component present in the input signal is removed by the resonator. This is an advantage of the new method because the DC usually poses a severe problem in all spatial filter approaches (Sanger 1988; Fleet et al. 1991).

Finally, disparity is determined from the phase difference between the resonator signals from both images. Phase comparison is achieved by multiplication of the two signals in the time domain and subsequent low-pass filtering (Fig. 1, Φ, *LP*).

Multiplication yields (Fig. 1, Φ):

$$\Phi(t) = y_l(t)y_r(t) = \begin{cases} g_{2f_0}(t) + \phi(t) & \text{if } t \geq t_r \\ 0 & \text{if } t < t_r \end{cases} \tag{6}$$

with:

$$g_{2f_0}(t) = \underbrace{a^2 e^{s_\infty(2t-t_l-t_r)} + a^{*2} e^{s^*_\infty(2t-t_l-t_r)}}_{\text{double frequency term}} \tag{7}$$

and

$$\phi(t) = \underbrace{|a|^2 e^{-s_\infty t_l - s^*_\infty t_r + (s_\infty + s^*_\infty)t} + |a|^2 e^{-s^*_\infty t_l - s_\infty t_r + (s^*_\infty + s_\infty)t}}_{\text{phase}}$$

$$= \underbrace{2|a|^2 \cos[(t_r - t_l)\mathrm{Im}(s_\infty)]}_{K} e^{\mathrm{Re}(s_\infty)(2t-t_r-t_l)} \tag{8}$$

The term $g_{2f_0}(t)$ reflects an oscillation with $2f_0$. In an implementation, it will be eliminated by low-pass filtering with low cut-off (Fig. 1, *LP*). The second part represents the phase $\phi(t)$ between the two signals and contains an exponential relaxation term and a constant term K, which encodes the true disparity.

$$K = \frac{Q^2}{2\pi^2 f_0^2 (4Q^2 - 1)} \cos[(t_r - t_l)\mathrm{Im}(s_\infty)] \tag{9}$$

The disparity, which is the spatial equivalent of $t_r - t_l$, can be computed by inverting (9) and is obtained immediately at the second contrast step (i.e., for $t = t_r$), after which the signal relaxes to zero. This relaxation behavior originating from the characteristic of the resonator assures temporal (viz. spatial) locality. Otherwise, only the average phase (viz. disparity) of each image line could be computed.

Like all other phase-based approaches, our algorithm is also subject to the so-called phase wrap-around problem. The periodic characteristic of the resonators limits the resolution of the system. This generic problem is reflected at the output of the system in (9) by the periodic behavior of the cosine. To avoid such ambivalencies, we restrict the argument of the cosine to: $0 < (t_r - t_l)\mathrm{Im}(s_\infty) < \pi$. From this we get: $0 < f_0 < \frac{1}{2}(t_r - t_l)^{-1}$; a constraint which is similar to that observed in the spatial filter (Gabor filter) approaches. The phase wrap-around problem disappears for $f_0 \to 0$ at the cost of low spatial resolution and an increasing noise susceptibility because of the shallow filter characteristic. As in the other spatial phase-based stereo-algorithms, our approach could also be used in a cascaded way, utilizing several filter-modules with different frequencies in order to address the phase wrap-around problem.

We obtain for the damping coefficient $Q > 1/2$, which means that the whole resonance may be restricted to about one half-cycle of the sine-wave. Given that disparity changes rarely exceed 5–10 pixels (empirical observation from publicly available stereo-image data), this restriction drastically limits the necessary computational effort in any implementation.

An analytical solution can also be obtained for other simple functional descriptions of disparity changes. In general, however, all disparity changes can be detected by such a system regardless of their shape as long as the frequency content of the change contains enough power at the resonance frequency.

The block diagram in Fig. 1 shows that this system can be easily implemented in analog or digital hardware. In particular, a few modern digital signal processors can be used to implement the individual filters, which are then coupled to a rather simple nearly-real-time processing system such as the one used to generate the data in the figure. In such a system, the disparity is determined continuously from the incoming data and the computational delay observed in the implementation (Fig. 1, ϕ) is constant. Its duration is mainly determined by the low-pass filter (Fig. 1, *LP*), and it is independent of the input image. In order to make this algorithm applicable, the output signal needs to be normalized to be independent of overall luminance variations. Such a normalization has been performed to obtain the results shown in Fig. 4 and 6.

Figure 2 shows what the signal originating from a single scan-line looks like at the different stages of the filtering process. The aperiodic brightness signal becomes transformed into a quasi-periodic signal at the resonator, where only a single frequency dominates (see spectra). The phase comparison (by multiplication) produces a signal with a DC and a 'double-frequency' component. Only the DC-component survives the low-pass filtering, and – as explained above – this DC signal represents the phase and, hence, the disparity.

3 Results

In the following, we shall compare our approach with several existing techniques. The next section gives some basic background about the algorithms used for comparison.

3.1 Other methods for disparity estimation

Several techniques have been proposed to recover depth information from epipolar line pairs. The classical approach uses a measure of similarity, cross-correlation for example, to find matching points in the two images composing the stereo pair. This technique selects one image of the stereo pair, for example the left one, as the

reference image. For each point of the reference image, the corresponding point is sought in the other image by searching for a maximum in the similarity measure along the corresponding epipolar line. To this end, the algorithm selects a rectangular window around a point in the reference image and computes its similarity measure with all the rectangular windows surrounding every point on the corresponding line in the second image. The point in the second image where the similarity measure has its maximum is considered the correct match.

This scheme, called 'area-based matching' (Haralick and Shapiro 1992) can be implemented in quite different ways, depending on the chosen similarity measure, on the algorithmic solution, and on the complexity of the model assumed for the disparity field. The similarity measures frequently used are sum of products, covariances, sum of squared differences, sum of absolute differences, and cross-correlation. The algorithmic solutions range from complete search to iterative least squares, simplex algorithms, and dynamic programming, highly depending on the a priori knowledge about the scene, the similarity measure, and the model of the disparity field. The model of the disparity field varies from simple translation (horizontal plane) to affinity (locally planar surface) to smooth (smooth surfaces without occlusions) or piecewise smooth (piecewise smooth, possibly with occlusions).

For comparison purposes, we implemented an area-based stereo algorithm that uses extensive search to identify the minimum at integer position, then the subpixel value of the minimum is computed via cubic interpolation of the similarity function. We produced disparity maps of a test scene with different similarity measures. The assumed model of the disparity field is that of a locally constant disparity. We denote the signal of each corresponding pair of scan lines as $f_R(x,y)$ and $f_L(x,y)$, where the subscript indicates that the scan-line comes from the right or the left image of the stereo pair. W_x and W_y define the size of the window in the x and y coordinates, respectively. Using this notation, the cross-correlation of point (x,y) with the disparity value d is defined as:

$$\mathrm{CC}(x,y,d) = \sum_{i=-W_x}^{W_x} \sum_{j=-W_y}^{W_y} f_L(x+i,y+j) \times f_R(x+i+d,y+j) \tag{10}$$

Plain cross-correlation is too sensitive to the local characteristics of the signal to be used in real applications. A better alternative is to use the zero-mean cross-correlation:

$$\mathrm{ZCC}(x,y,d) = \sum_{i=-W_x}^{W_x} \sum_{j=-W_y}^{W_y} (f_L(x+i,y+j) - \overline{f_L}) \times (f_R(x+i+d,y+j) - \overline{f_R}) \tag{11}$$

where $\overline{f_{R/L}}$ are the means of the signals in the windows. An even better alternative is the normalized cross-correlation:

$$\mathrm{NCC}(x,y,d) = \sum_{i=-W_x}^{W_x} \sum_{j=-W_y}^{W_y} \frac{f_L(x+i,y+j) f_R(x+i+d,y+j)}{\sigma_{f_L} \sigma_{f_R}} \tag{12}$$

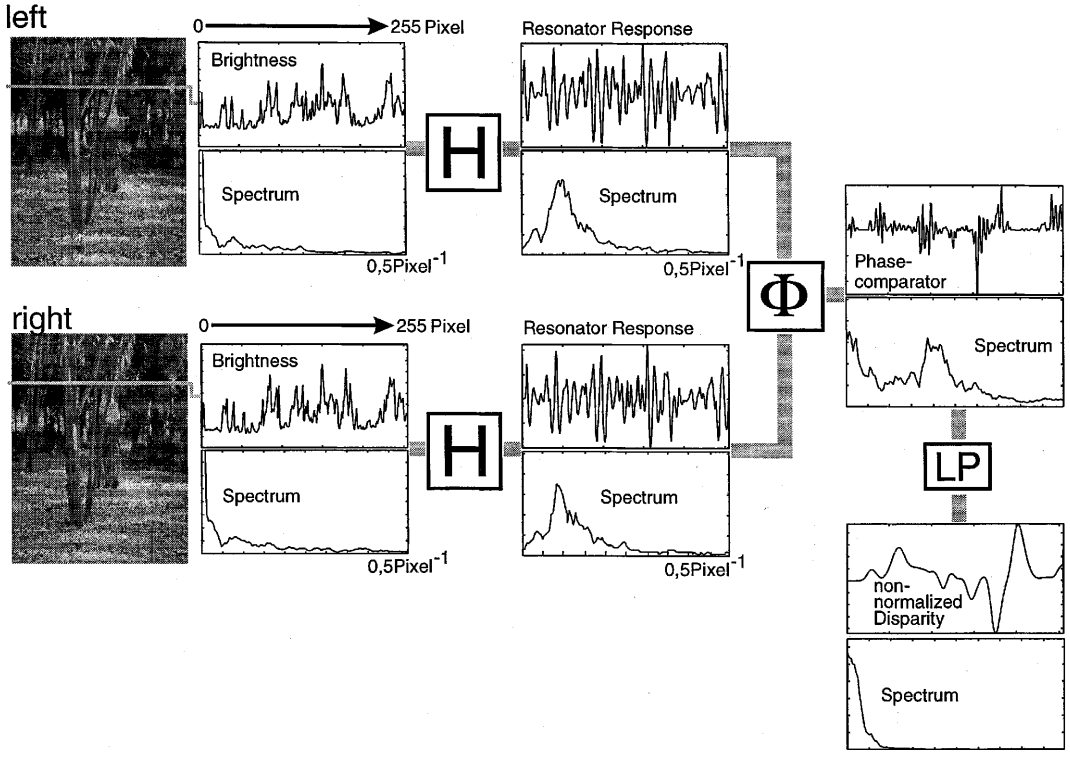

Fig. 2. Shape of the signals from a complete scan-line at different filtering stages and the power-spectra from these signals

where $\sigma_{f_{R/L}}$ are the standard deviations of the signals in the windows. Another approach is to search the *minimum* of the sum of squared differences between the two signals:

$$\mathrm{SSD}(x,y,d) = \sum_{i=-W_x}^{W_x}\sum_{j=-W_y}^{W_y}[f_{\mathrm{L}}(x+i,y+j) - f_{\mathrm{R}}(x+i+d,y+j)]^2 \quad (13)$$

As an alternative to correlation-based techniques, Sanger (1988) proposed the use of the phase difference between two local filter responses to compute the disparities of the different objects in the two stereo images. To achieve this, Gabor filters are commonly used. In the phase difference method (Sanger 1988; Fleet et al. 1991), disparity is computed from the phase difference between the convolutions of the two stereo images with local bandpass filters. Since the two signals, $f_{\mathrm{R}}(x)$ and $f_{\mathrm{L}}(x)$, are *locally* related by a shift $\delta(x_0)$, i.e., in the vicinity of each point x_0

$$f_{\mathrm{L}}(x+\delta(x_0)/2) \approx f_{\mathrm{R}}(x-\delta(x_0)/2) \quad (14)$$

the *local* k_0 Fourier components of $f_{\mathrm{L}}(x)$ and $f_{\mathrm{R}}(x)$:

$$\hat{f}_{\mathrm{L/R}}(k_0) = \int \exp(-\mathrm{i}k_0x) f_{\mathrm{L/R}}(x)dx = \rho(x)_{\mathrm{L/R}} \exp(-\mathrm{i}\phi(x)_{\mathrm{L/R}})$$

are related by a phase difference equal to $\Delta\phi(x) = \phi_L(x) - \phi_R(x) = k_0\delta$.

We can extract the local Fourier components by convolving the images with a Gabor filter:

$$F_{\mathrm{L/R}}(x,k_0) = \int G(x-y)\exp(\mathrm{i}k_0(x-y))f_{\mathrm{L/R}}(y)dy = \rho_{\mathrm{L/R}}(x)\exp(\mathrm{i}\psi_{\mathrm{L/R}}(x)) \quad (15)$$

where $G(x-y)$ is the Gaussian function and k_0 is the tuning frequency of the filter:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right)$$

As a function of the spatial position, the phase of the filter response, $\psi(x)$, has a quasi-*linear* behavior dictated by the center k_0:

$$\psi(x) \approx \psi'(x_0)(x-x_0) \approx k_0(x-x_0) \quad (16)$$

The local frequency, i.e., the derivative of the phase $\psi(x)$, is generally close to the value of the center frequency k_0. In fact, the Gabor filter is a bandpass filter around k_0.

In the Fleet et al. (1991) algorithm, the disparity is extracted from the phase difference, $\Delta\psi(x) = \psi_{\mathrm{L}}(x) - \psi_{\mathrm{R}}(x)$, by expanding $\Delta\psi(x)$ to the second order in δ:

$$\delta(x) \approx 2\frac{[\Delta\psi(x)]_{2\pi}}{\psi'_{\mathrm{L}}(x) + \psi'_{\mathrm{R}}(x)} \quad (17)$$

The phase is not defined when the amplitude vanishes, i.e., when $\rho(x) = 0$ (singularity). Around these singular points, the phase is very sensitive to spatial or scale variations. As a consequence, the approximation of (17) fails, and the calculation of disparity in the neighborhood of a singularity is unreliable. The neighborhoods of singular points can be detected by means of Fleet et al. (1991):

$$S(x) = \sigma\sqrt{(\psi' - k_0)^2 + \left(\frac{\rho'}{\rho}\right)^2} \le T_1 \quad (18)$$

$$\rho(x)/\rho^* > T_2 \quad (19)$$

where T_1 and T_2 are opportunely chosen constants, and ρ^* denotes the maximum value of the amplitude. The first term of (18) measures the difference between the peak frequency, k_0, and the local frequency, $\psi'(x)$, in relation to the width of the filter $1/\sigma$. The second term of (18) measures local amplitude variations with respect to the spatial width σ. The relation in (19) measures the 'energy' of the response. The result at point x is accepted only if the above relations are satisfied. Usually, T_2 is set to $\approx 5\%$, and $T_1 \approx 1.25$.

3.2 Comparison of the results

In Figs. 3 and 4, we show the disparity maps produced by six different techniques for disparity estimation. The *temporal resonance* (TR) technique used parameters $Q = 1.5$ and $f_0 = 0.08$. For the phase-based difference technique of Fleet and Jepson, we used two different Gabor filters: the first with a modulation period of 10 pixels and half-octave bandwidth (FJ10-0.5) and the second with a modulation period of 20 pixels and one-octave bandwidth (FJ20-1). All the correlation-based techniques, *normalized cross-correlation* (NCC), *zero-mean cross-correlation* (ZCC), and *sum of squared differences* (SSD), used a window size of 7×3 pixels with a disparity limit of ± 20 pixels.

The temporal resonance technique produces intermediate results compared with the other techniques, except on the uniform part of the source image. In these area there is no way to measure disparity, and the resonator response slowly fades. The filter-based techniques produce results in characteristic 'bands' centered

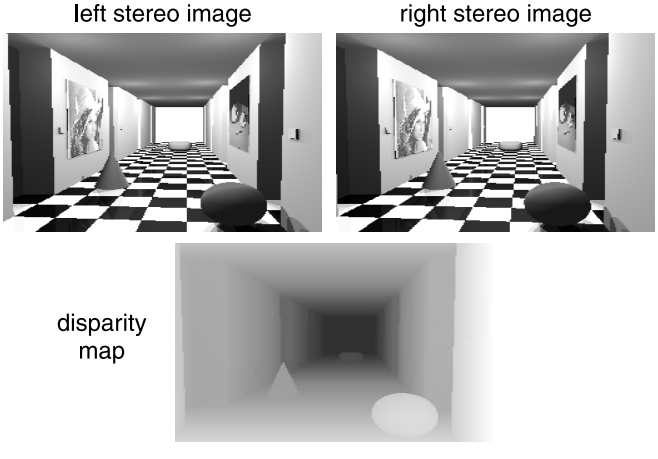

Fig. 3. The 'corridor' synthetic stereo-pair and its disparity map

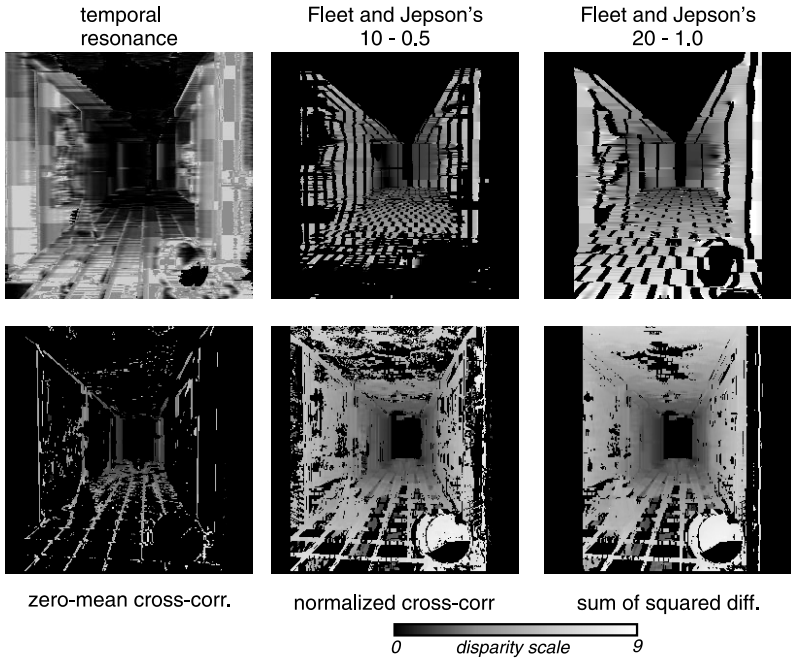

Fig. 4. The disparity maps produced by the different disparity estimation techniques

on the original image's edges. These bands are induced as a consequence of the spatial extent of the Gabor filter and have approximately the size of the filter's modulation period. The correlation-based techniques produce widely different results. In the case of ZCC, the differences in luminance cause the technique to find incorrect maxima in the correlation function, and thus most of the disparities are discarded in the verification phase. The resulting effect is that the disparity map is almost black. This problem is overcome by the normalization used by the NCC, that produces much better maps, visually very similar to the maps from the SSD.

In Fig. 5, we present a quantitative summary of the results from the different techniques. Two quantities are preponderant in characterizing the performance of disparity estimators: (1) *density*, i.e., the number of pixels where the algorithm is able to measure a disparity value, and (2) *precision*, i.e., the mean error that affects the measurements. By varying the parameters in the different algorithms, it is usually possible to trade density against precision or vice versa. Using as a test the synthetic image 'corridor' (Fig. 3), *temporal resonance* (TR) achieves the best results in density (91%) with an intermediate score in error (2.49 pixels). The comparison with NCC is interesting, a much slower technique that is still not able to beat it either in density or in precision. Fleet and Jepson's algorithm is very sensitive to the choice of the Gabor filter parameters (Cozzi et al. 1997), achieving very good precision (FJ20-1) or very poor precision (FJ10-.5) with nearly constant density around 40%. The ZCC produced the worse result in density (10%) with adequate precision (1.1 pixels). A good tradeoff is achieved by SSD, which gives the best precision (0.53 pixels) with a reasonable density (52%). The group of T. Kanade at CMU (Kanade et al. 1995) succeeded in producing a real-time implementation of the SSD, but this implementation is rather computationally expensive and, thus, requiring extensive hardware support.

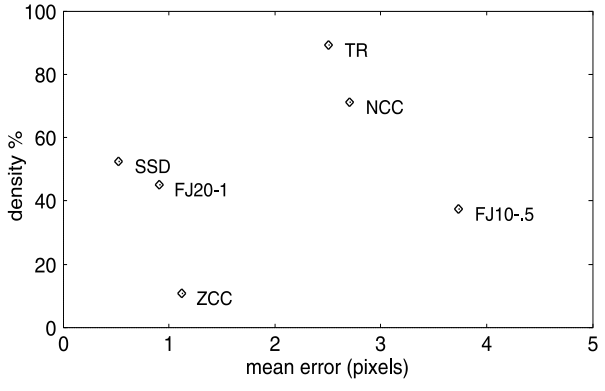

Fig. 5. Error and density of different disparity estimation techniques. Results for the 'corridor' synthetic test image. The labels have the following meaning: *TR* temporal resonance, *NCC* normalized cross-correlation, *ZCC* zero-mean cross-correlation, *SSD* sum of squared differences, *FJ* Fleet and Jepson, with two parameters for the Gabor filter: 20 and 10 pixels of tuning period, 1 and 0.5 octaves of bandwidth

Figure 6 shows an example of the performance of our algorithm tested with a real image pair. Disparity is retrieved with sufficient accuracy, but the map is more blurred than that obtained from the artificial scene. A comparison of this result with those obtained from the other algorithms (not shown) demonstrates that the performance of the temporal resonance algorithm falls well within the range of the other approaches.

4 Discussion

The theoretical framework presented here is based on the combination of computational principles found in the auditory and visual system of vertebrates. The convolution of the stereo-images with oscillating local

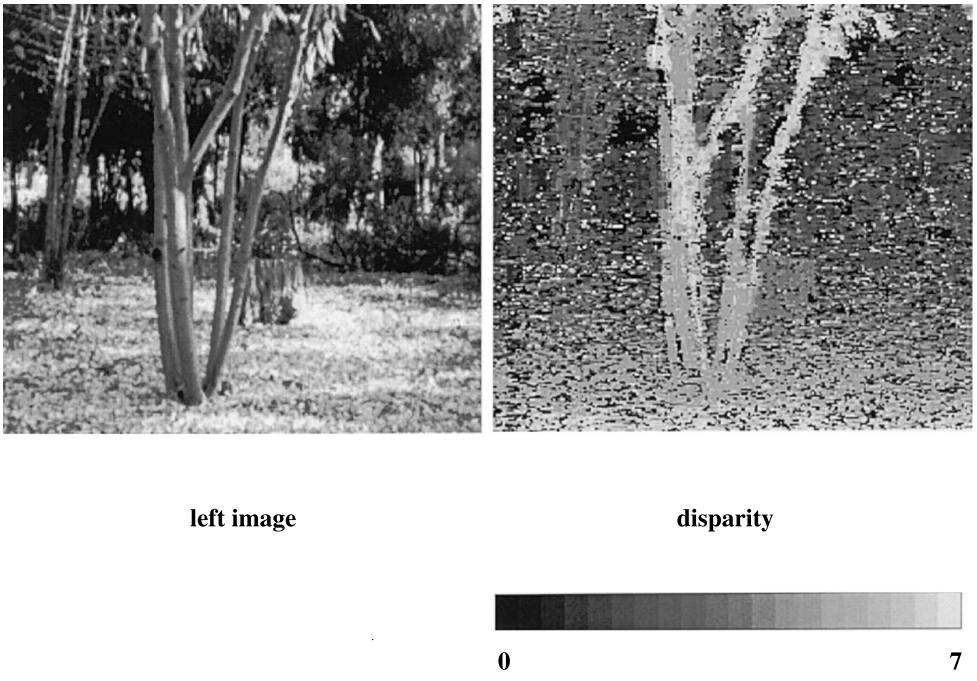

Fig. 6. Disparity map obtained by the temporal resonance algorithm from a stereo-image pair only the left image of which is shown. The disparity is coded with a gray scale (*bottom*), parameters were: $f_0 = 0.1443\,\text{pixel}^{-1}$, $Q = 1.0$

filters (Eq. 4) is a step quite commonly performed in the majority of technical approaches (Sanger 1988; Fleet et al. 1991) and reflects the response of cortical simple cells which have Gabor-like receptive fields (Daugman 1980; Marcelja 1980; Jones and Palmer 1987). It has been suggested that the evaluation of spatial phase differences from such receptive field responses could indeed be used to compute disparity in the brain (De Angelis et al. 1991). The dense coverage of the visual field by cells with many different receptive field sizes necessary to produce reliable depth maps does not pose a problem for the massive parallel architecture of the visual cortex. Technical systems, however, soon reach their limits with the tremendously high computational effort of such an architecture.

The one-dimensional structure of an auditory signal, on the other hand, allows us to determine the phase differences of two incoming sound waves sequentially by temporal correlation of both signals, and such a process probably takes place in the auditory cortex (Konishi and Sullivan 1986). The interpretation of image lines as sequential signals allows for a similar 'temporal' processing. The computational complexity of the temporal correlation involved, however, is reduced to simple multiplication and low-pass filtering as the consequence of the previously applied resonance filter. The transfer from the auditory to the visual domain worked well concerning the design of the novel algorithm. It should not be forgotten, however, that the algorithm shown here is rather unlikely to play any role in the visual system of the brain. After all, our visual cortex does not perform scan-line analysis.

On the other hand, the system is ideally suited for technical implementation in serial data acquisition systems, and it performs in nearly-real-time. More than that, the combination of spatial and temporal computational principles similar to those in the visual and auditory system generates a theoretical framework for causal stereoscopic depth processing in which the computational effort is strongly reduced. The comparison of the novel algorithm with other well-known techniques shows that intermediate results are obtained. Better results require rather complex algorithms, and real-time performance is then mostly prevented. Thus, our novel approach may be a good compromise in all those situations where real-time performance is necessary and a limited accuracy is sufficient. It may be possible to improve the performance of the temporal resonance algorithm by wiring up several algorithmic modules with different parameters in parallel. Such an architecture would still operate in real-time, and the results should be more correct. In particular, the so-called stereo-correspondence problem could also be addressed by combining modules. The correspondence problem always arises if the same features (here gray levels) occur more than once on an image scan-line. In that case, the match between left and right image becomes ambiguous. One could use very wide filters to avoid this problem, but these filters would almost always average over different disparity changes, leading to a wrong estimate. Thus, commonly used spatial approaches implement filter cascades of different widths and combine their results in order to achieve more accuracy and to reduce or eliminate the correspondence problem. To this end, we are currently investigating the theoretical background for such a combination of temporal resonance modules. It should be clear that the spatial resolution of all techniques which use filter cascades is usually reduced due to the limited spatial resolution of the filter with the lowest frequency. For this reason, Henkel (1994) has designed a more intelligent approach of combining different filters without changing their spatial frequency.

So far, our algorithm remains restricted to one dimension. Logically, an additional extension of the algorithm would be to try to combine the results from several scan-lines. Due to the continuity of objects, many times similar or identical disparity changes can be tracked over a certain vertical distance. Thus, it would make sense to combine the results from different scanlines obtained from our algorithm in order to exploit vertical disparity continuities. Our algorithm in itself ignores the second dimension, but this important source of information could be re-introduced afterwards by means of regularization techniques. Then, real-time performance would depend on the speed of the regularization method but seems still obtainable when using fast and simple techniques (e.g., averaging).

Acknowledgements. We are grateful to R. Opara for critical comments on the manuscript. The 'corridor' test image is courtesy of the Computer Vision Group of Prof. D. Fellner, Computer Science Dept., Bonn University, Germany. F.W. acknowledges the support of the Deutsche Forschungsgemeinschaft and the European Community ESPRIT 3, BRA 8305. A patent is pending for this system. In addition, preliminary results using multiple, cascaded filter modules can be inspected at: http://www.neurop.ruhr-uni-bochum.de/Real-Time-Stereo.

References

Cozzi A, Crespi B, Valentinotti F, Wörgötter F (1997) Performance of phase-based algorithms for disparity estimation. Machine Vis Appl 9:334–340

Daugman J (1980) Two-dimensional spectral analysis of cortical receptive field profiles. Vision Res 20:847–856

De Angelis G, Ohzawa I, Freeman R (1991) Depth is encoded in the visual cortex by a specialized receptive field structure. Nature 352:156–159

Fleet D, Jepson A, Jenkin M (1991) Phase-based disparity measurement. Comput Vision Graph Image Proc 53(2):198–210

Haralick RM, Shapiro LG (1992) Computer and robot vision, Vol 1. Addison Wesley, New York

Henkel RD (1994) Hierachical calculation of 3d-structure. (Technical report: Zentrum für Kognitionswissenschaften, Universität Bremen. http://axon.physik.uni-bremen.de/rdh/research/

Howard I, Rogers B (1995) Binocular vision and stereopsis. (Oxford Psychology Series no. 29) Oxford University Press, New York

Jones J, Palmer L (1987) An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol 58:1233–1258

Kanade T, Kano H, Kimura S (1995) Development of a video-rate stereo machine. In: Proc. of International Robotics and System Conference (IROS-95), Pittsburg

Konishi M, Sullivan W (1986) Neural map of interaural phase difference in the owl's brainstem. Proc Natl Acad Sci USA 83:8400–8404

Marcelja S (1980) Mathematical description of the responses of simple cortical cells. J Opt Soc Am 70:1297–1300

Sanger TD (1988) Stereo disparity computation using gabor filters. Biol Cybern 59:405–418

Wagner H, Frost B (1993) Disparity-sensitive cells in the owl have a characteristic disparity. Nature 364:796–757