Local Stereoscopic Depth Estimation Using Ocular Stripe Maps

Kai-Oliver Ludwig[*], *Heiko Neumann*[**], *and Bernd Neumann*

Universität Hamburg, FB Informatik, AB KOGS
Bodenstedtstr. 16, W-2000 Hamburg 50, Germany

Abstract. Visual information is represented in the primate visual cortex (area 17, layer 4B) in a peculiar structure of alternating bands of left and right eye dominance. Recently, a number of computational algorithms based on this ocular stripe map architecture have been proposed, from which we selected the cepstral filtering method of Y. Yeshurun & E.L. Schwartz [11] for fast disparity computation due to its simplicity and robustness. The algorithm has been implemented and analyzed. Some special deficiencies have been identified. The robustness against noise and image degradations such as rotation and scaling has been evaluated. We made several improvements to the algorithm. For real image data the cepstral filter behaves like a square autocorrelation of a bandpass filtered version of the original image. The discussed framework is now a reliable single-step method for local depth estimation.

Keywords: stereopsis, primary visual cortex, ocular stripe maps, cepstrum, local depth estimation.

1 Introduction

Multiframe analysis of images, such as stereopsis and time-varying image sequences, has been a primary focus of activities within the last decade of computational vision research. In both areas, the key problem has been identified as finding the correct correspondences of homologous image points. The so-called correspondence problem has not been solved to date to apply for general purpose vision tasks. For a review of relevant techniques for finding stereo correspondences, we refer to e.g. [3].

Finding stereo correspondence can be identified as a mathematically ill-posed problem, which has to be regularized utilizing constraints imposed on the possible solution (see e.g. [10]) . The majority of computational approaches is therefore formulated as finding a solution in a high dimensional search or optimization space by minimizing a functional which usually takes into account a data similarity term as well as a model term (e.g. for achieving smoothness) to regularize the solution (see e.g. [1, 10]). In order to avoid the complexity of most of the existing computational techniques, we investigated biological findings about architecture and mechanisms for seeing stereoscopic depth.

Due to the limited space of the current conference proceedings, this contribution does not cover all the topics we had to present. Neither does it provide you with the necessary context to be able to fully understand the presented facts. If required use [5, 6] and the references in there to get full background information.

---

[*] Email: ludwig@kogs26.informatik.uni-hamburg.de
[**] Email: neumann_h@rz.informatik.uni-hamburg.de

2 Biology

Biological Data Structures Support Efficient Computation. An alternative to the most commonly realized strategy in computational vision research ([8]) is to infer information processing capabilities from the identification of structural principles in the mammalian visual cortex (see e.g. [7]). These general principles include the discrete mapping of different sensory features like orientation, color, ocularity, depth or motion in 3D space to positions in a subspace of $\mathcal{R}^2$ ([4]). The "computational maps" discovered so far, have been postulated to optimally support computational mechanisms of different specificity.[3] Our computational model for local depth estimation is based on the subdivision of the cortex into ocular dominance stripes. Starting with the idea that a hypothetical disparity sensitive cell uses a local section of two neighboring stripes as input to compute local disparity, we can subdivide the original left and right image into local patches, whose size is chosen according to two major – in principle contradicting – constraints: increasing stability of estimate with increasing stripe width and, increasing accuracy with decreasing stripe width. Given the disparity at all single locations we obtain a disparity map from which a (relative) depth map can be easily inferred.

Usefulness of Stripe Maps for Depth Estimation. With reference to biological vision systems, we assume in this work a geometry in which the optical axes of the left and right image frames fixate a previously identified point in 3D space.[4] Then under these conditions of imaging geometry a circle is uniquely defined by the two optical centers of projection and the point of fixation. All points in space lying on this so-called *Vieth-Müller Circle* project onto the two retinae with zero disparity (Thales theorem). Due to the discrete width of the ocular stripes, not only projections of space locations with zero disparity can be fused. All 3D points with moderate negative (far field) or positive (near field) disparity within a psychophysically defined region also contribute to a fused image of varying depth due to the retinal shift of projection to retinal coordinates.[5] 3D spatial locations outside this area produce the well known phenomenon of double images.

3 Analysis, Evaluation, and Extensions

3.1 Analysis and Evaluation

To determine the values of the parameters of the technical model we have evaluated the relevant and sometimes diverging biological data from various sources to get a reasonable and consistent parameter setting. For a detailed discussion see [5]. Two corresponding local image patches extracted from the left and right image, respectively, can be arranged

---

[3] However, from the set of maps and principles given above, only the principles of retinotopy and ocular dominance stripe maps are fully established ([4]). The organization of alternating bands of ipsi- and contralateral eye dominance has been modeled recently as being the result of a structural transformation principle described via a non-linear mapping function ([7]). These so-called *ocular dominance columns* subdivide the whole area of the cortex (area 17, layer 4B) in alternating bands of ca. 0.5mm width.

[4] As a part of an active vision system, a fixating binocular head necessarily requires an attentional control module for the selection of appropriate fixation points and a module for the vergence movement to fixate the selected points. Proposals have been published how fixation points could be selected and how such a point may be tracked in time (see [5] for references).

[5] In case of idealized circular retinae the iso-disparity lines are circles of different radii with the horopter circle as one element of the set. In case of flat projection planes these iso-lines are conic sections (see [5] for a detailed discussion).

in a local neighborhood to form a single joint signal. This idea has been originally utilized in an algorithm proposed by Yeshurun & Schwartz [11] using rectangular patches butted against each other. If such a combined signal is filtered with the cepstrum[6], the filtered image contains a strong and sharp peak at a position which codes the disparity shift between the two original subsignals. This can be derived mathematically for the ideal case of a pure translational shift (see e.g. [5, 11]). Excluding some special cases – which will be named later in this paper – the disparity between the two subsignals can be obtained by simple maximum detection in the cepstral plane.

Using this method for computing disparities has several advantages. First, it is fast, because the disparities are computed in a single step without any iterations[7]. Second, due to the local and therefore independent computation of the disparities, parallelization is easy. Third, it is well-known from previous work, that the cepstrum is extremely insensitive to noise. We showed in a systematic evaluation ([5]) that the cepstral filter is insensitive to moderate image degradations due to rotation or scaling (6 degrees, 6%).

Geometry. It is reasonable to assume, that physical surfaces in the natural environment are piecewise smooth and can hence be approximated locally by their Taylor series expansion. We mathematically investigated the distortions in the disparity field when fixating planes and second order surfaces. For a given point in 3D space, let the left image coordinates be $\mathbf{l} = (x_L, y_L)$. Then the right image coordinates $\mathbf{r} = (x_R, y_R)$ can be computed in the first case to be:

$$x_R = \frac{a_1 x_L + a_2 y_L}{b_1 x_L + b_2 y_L + b_3} \quad \text{and} \quad y_R = \frac{c y_L}{b_1 x_L + b_2 y_L + b_3} \tag{1}$$

where a_i, b_i and c are constants with respect to a given stereo arrangement and local surface orientation (see [5] for further details on formulae and (graphical) results).

Evaluation. The cepstral filter as used in the literature with rectangular windowing functions suffers from some specific problems: If, for example, the double signal contains a single straight edge segment, then up to five additional maxima may appear in the cepstrum. In the case of varying illumination an additional peak at zero disparity may appear. These and other deficiencies can be overcome with different support functions.

3.2 Improvements and extensions

Other support functions for windowing. If rectangular support functions are butted against each other, information about the shape of the original subsignals is lost in the joint signal. Use of other window functions can prevent this information loss and improve signal properties. As a first example, we showed (see [5]) that the use of gaussian windows for the extraction of the local left and right image information produces a more easily identifiable maximum. Furthermore, the use of different support functions – as an approximation for a circular receptive field of a disparity cell – is also feasible and the results are better than those with standard rectangular support.

---

[6] The cepstrum (anagram of spectrum) is a well-known non-linear filter first used by Bogert et al. [2] for the detection of echo arrival times in 1D seismic signals. The cepstrum of a signal $g(\mathbf{x})$ is defined as $\mathcal{C}\text{epstrum}\{g(\mathbf{x})\} := \|\mathcal{F}\{log(\|\mathcal{F}\{g(\mathbf{x})\}\|^2)\}\|^2$. Due to its simplicity and noise robustness it has been widely used since then in various application areas from 1D speech processing to 2D image registration (see [6] for references).

[7] It has been shown in [9], that this filtering step could be done in 51ms when using special image processing hardware. With such short computation times it is feasible to use the presented cepstrum-based stereo segmentation approach in *active vision* systems for simple obstacle avoidance or object recognition tasks.

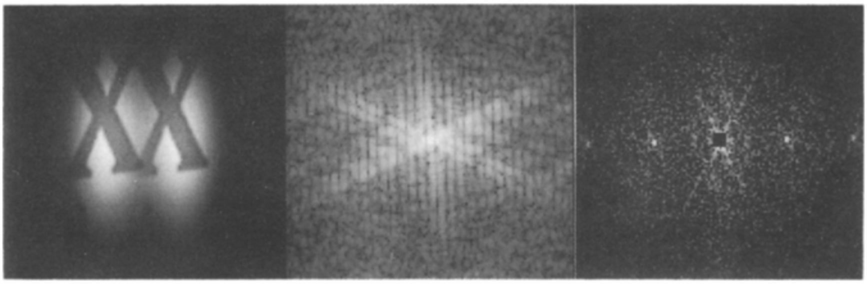

Fig. 1. Cepstrum with gaussian support functions. **Left:** Double signal $f(\mathbf{x})$ composed from data of left and right image at same (retinal) locations multiplied by a gaussian window and added with a fixed offset. **Center:** Amplitude spectrum. **Right:** $Cepstrum\{f(x,y)\}$. To enhance the visual impression $\log(\cdot)$ is displayed in the center and right image and a small region around the origin has been removed (only right).

The cepstral and autocorrelation. Based on a formula due to Olson & Coombs ([9]) the cepstrum can be written as autocorrelation preceded by an adaptive filtering step: $Cepstrum\{f\} = \|h \circ h\|^2$ with $h(x,y) = k_F(x,y) * f(x,y)$. For natural images we found that the prefilter, whose fourier transform is given by

$$K_F(u,v) = \frac{\sqrt{\log\left(\|F(u,v)\|^2\right)}}{\|F(u,v)\|} \quad \text{with} \quad F(u,v) := \mathcal{F}\{f(x,y)\} \tag{2}$$

computes mainly a bandpass filtered version of the original image with a narrow kernel, which is in contrast to the example given in ([9, p. 28]). On the other hand the cepstral filter is substantially better than autocorrelation applied to images appropriately filtered by LoG. It seems that the peculiar image-dependence of the prefilter kernel contains the main advantage of the cepstral filter.

4 Results, Conclusions and Prospects

We generated image pairs with a computer graphics visualization package to investigate the precision of the cepstral disparity estimates under precisely known conditions. The theoretical values given by (1) can be computed to ±1 pixel accuracy due to numerical instabilities and sampling effects. Fig. 2 gives an impression of the results on natural images.

Motivated by recent findings about the architecture of biological visual systems, we have investigated a method for how a binocular observer can recover local depth information with a single step computation avoiding the correspondence problem. In contrast to standard formulations of the stereopsis problem, this method needs neither regularization nor iterative computations to obtain the solution. It is a fast and reliable one-step method to determine depth locally around the point of fixation.

Current research topics include: For the technical aspects of the method, a mathematical analysis of support functions with good signal properties is necessary, since we currently investigated only box and gaussian shapes. In relation to this analysis, the bias introduced from the window shape as an error component in the estimation has to be evaluated. We also plan to investigate in greater detail the properties of the prefilter. For the incorporation of this local depth estimation technique in an active vision system, the problem of combining multiple depth maps has to be analyzed.

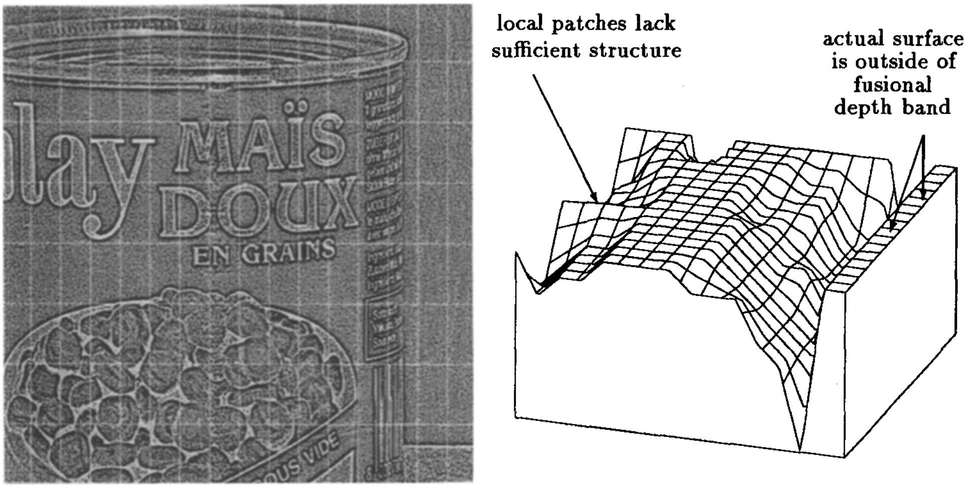

Fig. 2. Local depth map computed by the improved algorithm using equal gaussian window functions for left and right image with a previous LoG filtering step (The rectangles only outline the subdivision of the image). The image pair has been taken at a distance of 2 meters with stereo base length 7.00 cm using a precision adjusting device to produce the fixating arrangement. The (foveal) angle of extent is 200 minutes of arc. As can be observed the algorithm fails if one of the two indicated conditions hold (see arrows).

References

1. S.T. Barnard and M.A. Fischler: Computational and Biological Models of Stereo Vision. In *Proc. IU Workshop, Pittsburgh, PA, USA, September 11-13* (1990) 439–448
2. B.P. Bogert, M.J.R. Healy, and J.W. Tukey. The quefrency alanysis of time series for echoes: cepstrum, cross-cepstrum, and saphe cracking. In *Proceedings: Symposium on Time Series Analysis* (1963) 209–243
3. U.R. Dhond and J.K. Aggarwal. Structure from Stereo – A Review. *IEEE Trans. on Systems, Man, and Cybernetics*, **19(6)** (1989) 1489–1510
4. B.M. Dow. Nested maps in macaque monkey visual cortex. In K.N. Leibovic, editor, *The Science of Vision*, Springer, New York (1990) 84–124
5. K.-O. Ludwig. Untersuchung der Cepstrumtechnik zur Querdisparitätsbestimmung für die Tiefenschätzung bei fixierenden Stereokonfigurationen. Technical Report, Fachbereich Informatik, Universit"at Hamburg (1991)
6. K.-O. Ludwig, B. Neumann, and H. Neumann. Robust Estimation of Local Stereoscopic Depth. In *International Workshop on Robust Computer Vision (IWRCV '92), Bonn, Germany, October 9-12* (1992)
7. H.A. Mallot, W. von Seelen, and F. Giannakopoulos. Neural Mapping and Space-Variant Image Processing. *Neural Networks*, **3** (1990) 245-263
8. D. Marr. *Vision*. W.H. Freeman and Company, San Francisco (1982)
9. T.J. Olson and D.J. Coombs. Real-Time Vergence Control for Binocular Robots. Technical Report 348, Department of Computer Science, University of Rochester (1990)
10. T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, **317** (1985) 315–319
11. Y. Yeshurun and E.L. Schwartz. Neural Maps as Data Structures: Fast Segmentation of Binocular Images. In E.L. Schwartz, editor, *Computational Neuroscience*, Chap. 20, The MIT Press (1990) 256–266

This article was processed using the LaTeX macro package with ECCV92 style