A VISION FRONTEND WITH A NEW DISPARITY MODEL

João Rodrigues and J.M. Hans du Buf
University of Algarve, Faro, Portugal
jrodrig@ualg.pt dubuf@ualg.pt

Abstract

We are developing a frontend that is based on the image representation in the visual cortex and plausible processing schemes. This frontend consists of multi-scale line/edge and keypoint (vertex) detection, using models of simple, complex and end-stopped cells. This frontend is being extended by a new disparity model. Assuming that there is no neural inverse tangent operator, we do not exploit Gabor phase information. Instead, we directly use simple cell (Gabor) responses at positions where lines and edges are detected.

1 Introduction

During the last decade, the modeling of processes in the visual cortex has become a mature research topic. Mostly on the basis of Gabor quadrature filters as a model of simple cells, models of complex cells and end-stopped cells have been developed, e.g. (Heitger et al., 1992). In addition, models for line/edge detection (Deemter and du Buf, 1996) and bar and grating cells (Petkov and Kruizinga, 1997) have become available. Hence, it is now possible to develop a vision frontend that integrates all types of processing and that can be used to explore higher-level tasks like object recognition.

Because of the ocular dominance columns in the primary cortex (Hubel, 1995), which bring retinotopic, orientation-specific projections of the left and right eye closely together such that neural dendritic fields can cover both, we can assume that disparity estimation is already done in, or starts at, the first layers. However, we do not yet know *how* this is done (Qian, 1997). One approach in computer vision is based on the derivative of the local Gabor phase, also called the instantaneous frequency (Fleet et al., 1991). This method needs a coarse-to-fine scale stabilization and has problems in regions where the Gabor amplitude is very small, but it can provide a very rich and continuous disparity estimate. The Ohzawa-type model (Ohzawa et al., 1997) also employs quadrature Gabor filters, but it is more straightforward because disparity-sensitive complex cells are modeled by summing the squared responses of disparity-sensitive simple cells, the latter combining (positive) on and off Gabor responses of the left and right images. However, this model is still rather "academic" because it has not been tested with real images (the most complex example tested was a random dot stereogram, showing a very unstable disparity at the border, see (Qian, 1997)). It is likely that the Ohzawa-type model also needs a lot of postprocessing before it can be applied to real images.

Instead of experimenting with the Ohzawa-type model and the necessary stabilization, we decided to explore an alternative that is not based on phase nor amplitude summations. The basic idea is extremely simple: once we have line/edge detection, we have also access to the central, linear part of the Gabor responses, i.e. the sinusoidal or imaginary response in the case of lines and the sinusoidal-like or real response in the

case of an edge. Below we first introduce the basic keypoint and line/edge extraction with the necessary stabilizations, and then the new disparity model.

2 Lines, edges and keypoints

Line, edge and keypoint detection are based on the responses of simple, complex and end-stopped cells. Gabor filters provide a simple model of cortical simple cells. In the spatial domain they consist of a real cosine and an imaginary sine, both with a Gaussian envelope. Since all filtering is done in the frequency domain, we apply polar-separable transfer functions $G_{i,j}(f,\rho) = \exp\left(-(f-f_j)^2/2\sigma_{f_j}^2 - (\rho-\rho_i)^2/2\sigma_\rho^2\right)$, with $i = 0,\ldots N_\theta - 1$ and $j = 0,\ldots N_s - 1$, σ_f and σ_ρ being the bandwidths in the frequency and orientation, and N_s and N_θ the number of scales and orientations. We apply a linear scaling between $f_{\min}$ and $f_{\max}$, with the possibility of using many contiguous scales or a few noncontiguous scales with neighboring micro-scales, see below. We use $N_\theta = 8$.

In the spatial domain, the responses of even and odd simple cells, which correspond to the real and imaginary parts of the Gabor filters, are denoted by $R^E_{s,i}(x,y)$ and $R^O_{s,i}(x,y)$, s being the scale and i the orientation (i.e. $\theta_i \perp \rho_i$ and $\theta_i = i\pi/(N_\theta - 1)$). In order to simplify the notation, and because the same processing is done at all scales, we drop the subscript s. The responses of complex cells are modelled by the modulus $C_i(x,y) = [\{R_i^E(x,y)\}^2 + \{R_i^O(x,y)\}^2]^{\frac{1}{2}}$. There are two types of end-stopped cells (Heitger et al., 1992; Würtz and Lourens, 1997) , i.e. single (S) and double (D). Introducing $\mathcal{C}_i = \cos\theta_i$ and $\mathcal{S}_i = \sin\theta_i$, these cells are modelled by

$$S_i(x,y) = [C_i(x + d\mathcal{S}_i, y - d\mathcal{C}_i) - C_i(x - d\mathcal{S}_i, y + d\mathcal{C}_i)]^+ ,$$

$$D_i(x,y) = \left[C_i(x,y) - \frac{1}{2}C_i(x + 2d\mathcal{S}_i, y - 2d\mathcal{C}_i) - \frac{1}{2}C_i(x - 2d\mathcal{S}_i, y + 2d\mathcal{C}_i)\right]^+ ,$$

in which $[\cdot]^+$ denotes the suppression of negative values. The distance d is scaled linearly with the filter scale s, i.e. $d = 0.6s$. All end-stopped responses along straight lines and edges need to be suppressed, for which we use tangential (T) and radial (R) inhibition:

$$I^T(x,y) = \sum_{i=0}^{2N_\theta - 1} [-C_{i \bmod N_\theta}(x,y) + C_{i \bmod N_\theta}(x + d\mathcal{C}_i, y + d\mathcal{S}_i)]^+$$

$$I^R(x,y) = \sum_{i=0}^{2N_\theta - 1} \left[C_{i \bmod N_\theta}(x,y) - 4 \cdot C_{(i+N_\theta/2) \bmod N_\theta}(x + \frac{d}{2}\mathcal{C}_i, y + \frac{d}{2}\mathcal{S}_i)\right]^+ ,$$

where $(i + N_\theta/2) \bmod N_\theta \perp i \bmod N_\theta$. Instead of applying the inhibition to individual end-stopped cells (Heitger et al., 1992; Würtz and Lourens, 1997), slightly better results are obtained by applying $I = I^T + I^R$ once to the pooled activity:

$$K^S(x,y) = \sum_{i=0}^{N_\theta - 1} S_i(x,y) - gI(x,y) \quad \text{and} \quad K^D(x,y) = \sum_{i=0}^{N_\theta - 1} D_i(x,y) - gI(x,y),$$

with $g \approx 0.4$, after which the keypoint map is obtained: $K(x,y) = \max\{K^S(x,y), K^D(x,y)\}$, i.e. at each filter scale s: $K_s(x,y)$.

Postprocessing of $K(x,y)$ is necessary because our goal is to obtain a clean, single-pixel keypoint map with no spurious points. In addition, for pattern recognition applications we want to classify the keypoints according to the underlying vertex structure, i.e. K, L, T, + etc. Postprocessing is done in six steps, at each scale, after which different scales are combined by considering, around each scale, a small scale interval: 4 *micro*-scales, i.e. two scales slightly finer and two slightly coarser than the actual scale. In the case of the smallest (largest) scale four coarser (finer) scales are applied. Only keypoints which are consistent over 3 neighboring micro-scales are accepted.

Line and edge detection is based on simple cells (Deemter and du Buf, 1996). A positive line is detected where R^E shows a local maximum and R^O shows a zero crossing. In the case of an edge the even and odd responses must be swapped. This gives 4 possibilities for positive and negative events: local maxima/minima plus zero crossings. We also combine the responses of simple and complex cells, i.e. the simple cells serve to detect positions and event types, whereas the complex cells are used to increase the confidence. Since the use of Gabor modulus (complex cells) implies a loss of precision at vertices (du Buf, 1993) we increase precision by considering multiple scales. Our improved detection scheme consists of 5 processing steps for each scale, after which the coherence is improved by checking 4 neighboring micro-scales.

3 Disparity estimation

Our disparity model is based on the central, linear part of the Gabor responses, i.e. the sinusoidal part with $\sin x \approx x, |x| < \pi/4$. Assuming ideal events, i.e. lines with a Dirac profile and edges with a Heaviside step profile, or nonideal ones obtained by Gaussian filtering, and complex Gabor filters with the same orientation, the responses are (scaled) Gabor functions and complex errorfunctions. The latter can be approximated by scaled Gabor functions (du Buf, 1993). In other words, both line and edge responses are essentially scaled Gabor functions. One step in line/edge detection consists of checking the Gabor response $R_i^O(x,y)$ (the odd, imaginary part in the case of a line), or $R_i^E(x,y)$ (the even, real (!) part in the case of an edge) for a zero crossing on $\pm\lambda/4$. Here, for disparity, we apply the same event detection steps to two images, left and right. In the case of the *left* image, we (1) check the existence of an event of the same type in the *right* image on $\pm\lambda/8$, and (2) if so, we take the value of $\pm R^O$ or $\pm R^E$ of the *right* image at the event (zero crossing) position in the *left* image. The sign depends on the event polarity and, in order to obtain values which do not depend on the event amplitude, $\pm R^O$ or $\pm R^E$ is divided by the modulus (complex cell response) of the *left* image, which is maximum at the event position. After this normalization yet another one is applied: the response is divided by the scale s of the filter. Hence, the slope of the linear response part will not depend on the event amplitude nor on the filter scale, i.e. disparity estimates obtained at different scales will be the same. The same processing can be done in the case of the right image, by exchanging *left* and *right*. Of course, the disparity estimates need to be calibrated once using real data, like the way babies need to learn in the first months.

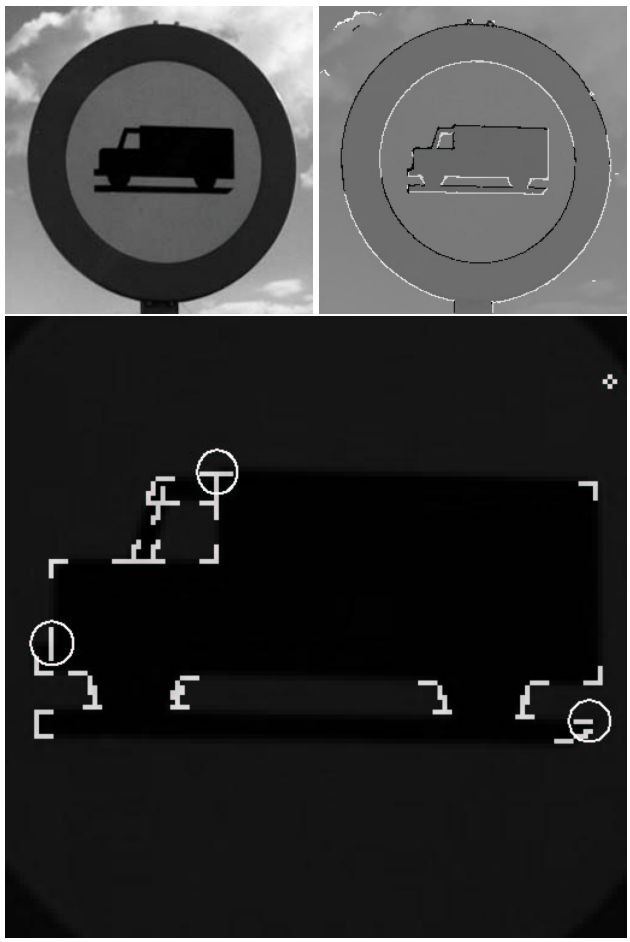

Figure 1: Top: sign2 image and line/edge detection. Bottom: keypoints detected with event directions, after multi-scale stabilization (zoomed).

One problem we encountered were small fluctuations of the disparity estimates, especially at the finest scales. These are due to the fact that we need to work at discrete pixel positions, and the maximum of the modulus used in the first normalization is therefore not the theoretical maximum. We solved this by averaging disparity estimates over neighboring scales.

4 Results and discussion

Figure 1 shows results of line/edge and keypoint detection. Single- and multi-scale stabilization has eliminated many spurious keypoints, one of which is shown by the small diamond of 4 pixels (bottom image). All keypoints of the van have been detected,

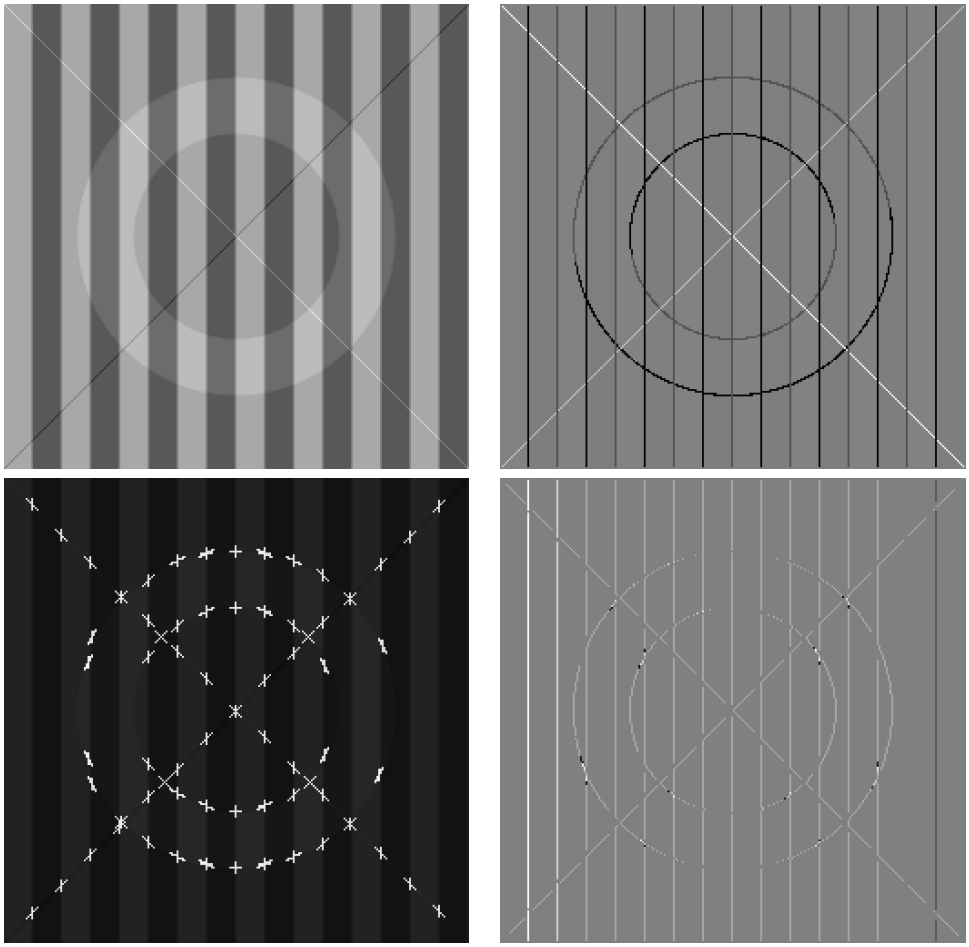

Figure 2: Ledge image (top-left), line/edge detection (top-right), keypoints (bottom-left) and disparity (bottom-right). See text.

but three directions are still missing (encircled). Here the structures have a size of 2 to 4 pixels: we are at the limit of what can be achieved by using Gabor wavelets.

Figure 2 shows, apart from line/edge and keypoint detection, disparity estimation (bottom-right). Event types (top-right) are coded by different gray levels, and distance in z (bottom-right) as well. Disparity was obtained by shifting left, in one image of a pair, the first vertical edge 3 pixels, the following edge 2, the next edges 1 pixel. The second-last edge was not changed, and the last was shifted right. The lines and ring were shifted left 1 pixel. As can be seen, disparity estimates are proportional to the shifts, but there are still problems at keypoints. In addition, a comparison with traditional methods and experiments with real images showed that the interval $\pm\lambda/8$ of the filters is too small, even of the biggest filters. The reason is that these filters are the smallest ones in the frequency domain. This is being solved by creating bigger filters,

doing the filtering by convolution in the spatial domain. The main conclusion is that it will be possible to create a sort of "wireframe" representation in which lines and edged are tagged with disparity, simplifying 3D object recognition. The same might occur in our visual cortex.

Acknowledgement Part of this research is supported by ISR/IST in Lisbon, FCT Programa Operacional Sociedade de Informação (POSI) in the framework QCA III.

References

Deemter, J. v. and du Buf, J. (1996). Simultaneous detection of lines and edges using compound Gabor filters. *Int. J. of Pattern Recognition and Artificial Intelligence*, 14(6):757–777.

du Buf, J. (1993). Responses of simple cells: events, interferences, and ambiguities. *Biological Cybernetics*, 68:321–333.

Fleet, D. J., Jepson, A. D., and Jenkin, M. R. (1991). Phase-based disparity measurement. *CVGIP: Image Understanding*, 53:198–210.

Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., and Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vis. Res.*, 32:963–981.

Hubel, D. H. (1995). *Eye, brain and vision.* Scientific American Library.

Ohzawa, I., DeAngelis, G. C., and Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *J. Neurophysiol.*, 18(77):2879–2909.

Petkov, N. and Kruizinga, P. (1997). Computational models of visual neurons specialised in the detection of periodic and aperiodic visual stimuli. *Biological Cybernetics*, 76(2):83–96.

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, 18:359–368.

Würtz, R. and Lourens, T. (1997). Corner detection in color images by multiscale combination of end-stopped cortical cells. *Artificial Neural Networks – ICANN'97*, Springer LNCS Vol. 1327:901–906.