AUTHORS:

Niklas Lüdtke
Department of Computer Science
University of York,
Heslington, York YO10 5DD, United Kingdom
e-mail: niklu@cs.york.ac.uk
phone: +44-1904-432774
fax : +44-1904-432767

Dr. Richard C. Wilson
Department of Computer Science
University of York,
Heslington, York YO10 5DD, United Kingdom
e-mail: wilson@cs.york.ac.uk
phone: +44-1904-434726
fax : +44-1904-432767

Prof. Edwin R. Hancock
Department of Computer Science
University of York,
Heslington, York YO10 5DD, United Kingdom
e-mail: erh@cs.york.ac.uk
phone: +44-1904-433374
fax : +44-1904-432767

Population Coding and the Detection of Visual Stimuli with Multiple Orientations

Niklas Lüdtke, Richard C. Wilson and Edwin R. Hancock

Department of Computer Science, University of York, York, YO10 5DD, UK

{*niklu, wilson, erh*}@*cs.york.ac.uk*

Abstract

We address the problem of detection and representation of stimuli with multiple orientations within the framework of a biologically motivated computer vision system for contour detection based on population coding. Due to the similarity between the receptive field properties of striate cortical neurons and oriented filters applied in image processing, there is also a close relation between artificial orientation estimation and models of orientation selectivity in computational neuroscience. Like a cortical hypercolumn, a bank of Gabor filters constitutes a population code of local stimulus orientation that allows a read-out of the encoded information by techniques such as the population vector or more refined probabilistic decoding strategies.

We present a probabilistic population decoding model of stimulus orientation with Gabor filters based on the analytically derived orientation tuning function and a parametric mixture model of the filter responses in the presence of stimuli with single or multiple orientations. A probability density function of local stimulus orientation is extracted through an EM-based parameter estimation procedure. The resultant probability density of local orientation captures not only angular information at edges, corners and occlusion points but also describes the certainty of the orientation estimate.

1 Introduction

Population coding has emerged as an essential paradigm in neurobiology and is increasingly studied among theorists in the computational neuroscience community. Georgopoulos and co-workers [2] introduced the concept of the *population vector* and demonstrated how the direction of arm movements could be decoded from neural firing rates in the motor cortex of monkeys. In the vision domain, population vector coding has also been linked to orientation estimation [5]. Recently, population coding has been extended to extracting entire probability densities from the ensemble activities [7]. This opens up the possibility of a more "holistic" processing of the information encoded in a population. Unlike "reductionist" methods, which aim to extract single quantities, the probabilistic approach preserves the distributed nature of the code and provides additional information about the *certainty* of the encoded variable(s). Moreover, the probability distribution can be multi-modal, signalling multiplicity in the encoded variable.

Two-dimensional stimuli with multiple orientations create multimodal response activity profiles in the hypercolumnar ensemble. How such stimuli are processed in the brain is currently debated, and experimental data is rare. Carandini and Ringach [1] presented a recurrent model of orientation selectivity, where excitatory and inhibitory intra-columnar connections are modelled by a "center-surround" weighting function operating in the orientation domain. However, the model exhibits peculiar responses in the presence of stimuli containing multiple orientations. For instance, if the stimulus contains two orientations that are separated by less than 45° the model hypercolumn is unable to veridically encode the orientations. Zemel and Pillow [8] developed a more refined recurrent population coding model based on an optimized weighting function, yielding more plausible results and an enhanced capability of resolving and representing multiple orientations.

In this paper probabilistic population coding is applied to a computer vision task, namely the detection and representation of local edge orientation with Gabor filters. The ability to represent ambiguous inputs is used to extract multiple orientations in corner points and T-junctions, where classical edge detectors usually fail. Though our approach is machine vision oriented, there are a number of common issues in artificial and biological orientation selection. Like a striate cortical hypercolumn, a bank of Gabor filters constitutes a population code of local stimulus orientation that allows a read-out of the encoded information by techniques such as the population vector [2] or more sophisticated probabilistic decoding strategies [7].

The crucial assumption in probabilistic population coding of multiple stimuli is that the ensemble activity profile is a linear superposition of the responses to individual stimulus components. In the conclusion we discuss some of the criticism levelled at purely linear processing in the computer vision literature and potential implications of non-linear processing strategies suggested therein for population coding.

2 Orientation Tuning of Gabor Filters

The tuning function of an odd-symmetric Gabor filter can be derived analytically for a sinusoidal grating of arbitrary orientation. The result is of general relevance since any fully anisotropic input can be expanded into a Fourier series of sinusoids with different wavelengths but equal orientation. Real images are likely to contain edges subject to some degree of blur that therefore have a dominant spatial ground frequency in their spectrum, while higher frequency components are comparatively weak. For curved edges the aforementioned still holds approximately within the effective range of the filter mask. The characteristic spatial frequency of the edge structure at a particular location influences the tuning width of the filter responses. For practical reasons the corresponding wavelength λ_s (in pixels) will be used. We consider a sinusoid of orientation θ

$$S(x,y) = \sin\left[k_s\left(x\cos\theta + y\sin\theta\right)\right]\ ; \quad k_s = 2\pi/\lambda_s . \tag{1}$$

For simplicity we choose a Gabor filter with aspect ratio one (radial symmetry), wavelength λ_f, vertical preferred direction and define analogously $k_f = 2\pi/\lambda_f$:

$$\mathcal{G}_{odd}(x,y) = \exp\left(-\tfrac{x^2+y^2}{2\sigma^2}\right)\sin(k_f x)\,. \tag{2}$$

The orientation tuning function $f(\theta)$ is the convolution of the filter with the sinusoid at the origin $(0,0)$.

$$f(\theta) = (\mathcal{G}_{odd} * S) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \mathcal{G}_{odd}(x,y)S(x,y,\theta)\,dx\,dy \ \propto\ \sinh\Big(\underbrace{k_f k_s \sigma^2}_{\kappa_0}\cos\theta\Big) \tag{3}$$

The parameter κ_0 is a so-called concentration parameter. Its reciprocal value is related to the angular variance and therefore controls the orientation tuning width. κ_0 depends on the two known filter properties $k_f = 2\pi/\lambda_f$ and σ as well as the unknown quantity $k_s = 2\pi/\lambda_s$. In the following it will be assumed that the tuning functions for all filters are identical apart from an angular shift ψ indicating the preferred orientation:

$$f(\theta;\, \psi) = C \sinh[\kappa_0 \cos(\theta - \psi)] \,. \tag{4}$$

Here $C = 1/\sinh(\kappa_0)$ is a normalization constant, so that $f(\theta;\, \psi) \in [-1, 1]$.

3 Population coding in the "Gabor-hypercolumn"

A Gabor filter bank is an ensemble of linear orientation selective units. Therefore, the *principle of superposition* holds: the response profile for complex intensity structure, such as in corner points, where several edges coincide, is a linear combination of the response activities for the individual edge components. This assumption is crucial for the following derivations.

While a straight edge has a single well defined orientation, problems occur when either the edge is curved or multiple orientations exist within the filter mask. In addition, noise in the image can create uncertainties even in the angle of a straight edge. We can accommodate all these possibilities by adopting a stochastic model of edge orientation where the edge angle θ is governed by probability density function $p(\theta)$. Curvature is signalled as a range of possible angles.

3.1 Expectation values of filter responses

Let ψ be the preferred orientation of a filter. Assuming that superposition holds (cf. section 3), the expectation value of the response profile is given by the convolution of the tuning function $f(\theta)$ and the pdf of the the stimulus orientation $p(\theta)$ [7]:

$$\bar{r}(\psi) = \int_0^{2\pi} f(\theta - \psi) p(\theta)\; d\theta \,. \tag{5}$$

This equation describes the process of encoding of the pdf $p(\theta)$ in the expected response function $\bar{r}$ which is continuous in ψ. The average ensemble activities (the average response profile) are a sample of this function for a discrete set of filter orientations ψ_ν.

Thus the decoding of the pdf $p(\theta)$ is a *deconvolution*, which is an ill-posed problem and likely to require some kind of regularization. In the following sections a parametric model of the expected filter responses is derived based on the orientation tuning function and a mixture model of $p(\theta)$. Here the regularization is *implicitly* achieved through the parametric approach. In the following we will derive a parametric model of the filter responses based on the orientation tuning function and a mixture model of $p(\theta)$.

3.2 A mixture model of local orientation

The encoding equation (5) is general in the sense that $p(\theta)$ can take any form and is not restricted to unimodality. We now choose a particular type for the pdf of edge orientation which we expect to see in the image. The probability density of the stimulus orientation is modelled as a mixture of von Mises distributions[4], where each principle edge orientation is represented by a mixture component with a mean value $\bar{\theta}_i$ and a concentration parameter κ_i:

$$p(\theta) = \frac{1}{2\pi} \sum_{i=1}^{m} \frac{P(i)}{I_0(\kappa_i)} e^{\kappa_i \cos(\theta - \bar{\theta}_i)}, \quad \text{with} \sum_{i=1}^{m} P(i) = 1\,. \tag{6}$$

Here I_0 is the modified Bessel function of first kind and order zero and the term $1/2\pi I_0(\kappa_i)$ serves as a normalization factor of the i-th mixture component. Eqn.(6) can be considered a circular analogue of the Gaussian mixture density. The κ_i correspond to the $1/\sigma_i$ and the $\bar{\theta}_i$ to the μ_i in a Gaussian mixture density. The $P(i)$ are the mixing coefficients. The number of mixture components, m, will be limited to two or at most three, which incorporates the essential cases of multiple edge orientation, i.e., corners and junctions.

4 The probability distribution of responses

In addition to eqn.(5) the expected response profile can also be thought of as an average over the filter responses themselves. Let $p(r;\psi)$ be the probability density over the response value of a filter of preferred orientation ψ. Then the expected (continuous) response profile is given by

$$\bar{r}(\psi) = \int_{-1}^{1} r\, p(r;\psi)\, dr\,. \tag{7}$$

The filter responses r_ν obtained at a certain location in an image are instances of stochastic variables even though the filtering is per se a deterministic operation. The randomness of the responses is created by the stochastic nature of the input variable alone. This is different from standard biological models of population coding in which neural firing rates are usually considered random variables following Poisson statistics [7].

4.0.1 Maximum likelihood estimation

One way of estimating the parameters of the mixture model $p(\theta)$ is through a maximum likelihood estimation. Let Θ denote a set of mixture parameters. It is essential to know the likelihood of the individual filter responses given their preferred orientations ψ_ν and the parameters Θ and κ_0, i.e. $p(r_\nu;\psi_\nu,\Theta,\kappa_0)$ in order to calculate the total likelihood of a given response profile.

$$\mathcal{L}\{r_1 \ldots r_n |\, \Theta, \kappa_0\} = \prod_{\nu=1}^{n} p(r_\nu;\psi_\nu,\Theta,\kappa_0)$$

The likelihood $\mathcal{L}$ depends on the parameters of the mixture pdf $p(\theta)$ plus the parameter κ_0 specifying the tuning function. Maximum likelihood estimation of the mixture parameters can be performed using standard techniques such as the EM algorithm. The remaining tuning parameter κ_0 can be obtained from previous measurement(s) using test stimuli (straight edges).

The task is now to find the pdf of a *function of a random variable* since the filters transform the edge orientation θ via their tuning function given by (4). There is insufficient space to fully derive the resulting distribution, so we limit ourselves to quoting the result:

$$\begin{aligned} p(r;\psi) \quad = \quad & \frac{1}{\pi C \kappa_0} \sum_{i=1}^{m} \frac{P(i)}{I_0(\kappa_i)} \frac{\exp\left[\frac{\kappa_i}{\kappa_0} \sinh^{-1}\left(\frac{r}{C}\right) \cos(\psi - \bar{\theta}_i)\right]}{\sqrt{1 + \left(\frac{r}{C}\right)^2} \sqrt{1 - \left[\frac{1}{\kappa_0} \sinh^{-1}\left(\frac{r}{C}\right)\right]^2}} \cdot \\ & \cosh\left(\kappa_i \sqrt{1 - \left[\frac{1}{\kappa_0} \sinh^{-1}\left(\frac{r}{C}\right)\right]^2} \sin(\psi - \bar{\theta}_i)\right) \end{aligned} \tag{8}$$

This density is again a mixture model. It is then possible to estimate the model parameters with the expectation maximization (EM) algorithm. Due to the complexity of the response pdf the resultant update equations for each mixture component form a transcendental system for the pair of parameters $(\bar{\theta}_i; \kappa_i)$. However, one can obtain closed form approximations with good accuracy.

5 Measuring certainty

The probabilistic approach not only yields an estimate for the different edge orientations present in the neighbourhood of the considered point (x, y) but also provides information about the certainty of these measurements through the concentration parameters, the κ_i. Zemel and colleagues [9] proposed the *resultant length* [4], which is given by ($\bar{\theta}$ is the mean direction, $E\{.\}$ denotes the expectation value)

$$\rho = E\{\cos(\theta - \bar{\theta})\} = \frac{I_1(\kappa)}{I_0(\kappa)} \tag{9}$$

as a measure of certainty within the framework of the "directional-unit Boltzmann machine". We argue that for practical purposes it is more suitable to use the Kullback-Leibler divergence of a given mixture component $p_i(\theta)$ from the distribution of maximum entropy, i.e. the uniform distribution $q(\theta) = 1/2\pi$.

$$K(p_i, q) = \int_0^{2\pi} p_i(\theta) \ln\left[\frac{p_i(\theta)}{q(\theta)}\right] d\theta = \frac{I_1(\kappa)}{I_0(\kappa)}\kappa - \ln[I_0(\kappa)] \tag{10}$$

This quantity, also called *relative entropy*, is, unlike the entropy itself, always positive. For convenience the relative entropy can be normalized through a sigmoid function $g(x) = \frac{2}{1+e^{-x}} - 1$ which yields:

$$\gamma = g[K(p_i, \tfrac{1}{2\pi})] = \frac{2}{1 + \frac{\exp(-\kappa I_1(\kappa)/I_0(\kappa))}{I_0(\kappa)}} - 1\,. \tag{11}$$

A comparison of the certainty measures is shown in Fig. 1 (left). Our measure allows better distinction of certainties for concentration parameters of middle range.

6 Results

Figure 1 (right) shows an image containing three objects which generate a number of different edge configurations including straight edges, curved edges, corners and T-junctions.

The EM-algorithm described in the previous sections was applied to a number of key points in the image. For the purpose of better visualization, the pdf is displayed as a polar plot; the angle represents θ and the radius $p(\theta)$. Each von Mises mixture component produces a club-shaped plot oriented in the same way as the the corresponding detected orientation. The *direction* of the "club" depends on the sign of the contrast. In this case, a bright-dark transition (counter clock-wise) yields an angle $\theta \in [0°, 180°)$ and a dark-bright transition an angle $\geq 180°$.

Figure 2 (top left) shows the original filter bank response from point (1). The pdf of the orientation angle, which results from the EM-algorithm, is shown in Figure 2 (bottom left). Point (1) is a corner point; as expected, two well pronounced and roughly perpendicular components are found (certainties $\gamma_1 = 0.73$ and $\gamma_2 = 0.75$). The second set of results (Figures 2, middle) comes from point (2), which is a straight edge. The EM-algorithm finds a strong component (certainty $\gamma = 0.8$) in the direction of the edge. Finally, Figures 2 (top right and bottom right) stem from point (5), which is a T-junction created by an occluding edge. This edge is also curved, but because the edge is occluding there are only two principle edge directions. Consequently, the pdf contains a large component (certainty $\gamma = 0.74$) for the curved edge and a smaller component with a slightly higher certainty for the weaker secondary edge (certainty $\gamma = 0.75$).

7 Conclusion and Discussion

We have presented a framework that applies the concept of probabilistic population coding to local edge orientation estimation with a bank of odd-symmetric Gabor filters. Based on the assumption that local edge orientation follows a von Mises distribution, edges as well as points of multiple orientation, such as corner points and T-junctions, can be modelled by a von Mises mixture distribution. Given the filter responses at a particular location, the parameters of this angular distribution are estimated by means of an EM-algorithm.

The linearity of the filters is essential for the validity of the mixture model, since it ensures that responses to stimuli with multiple orientations follow the principle of linear superposition. Despite successful interpretations of physiological data in multiple motion perception [6], the issue of superposition in biological vision is clearly more debatable and the amount of available physiological data is insufficient.

There are signal-theoretic arguments questioning the efficiency of linear filters in early vision. For instance, Barth and Zetzsche [10] demonstrated that a linear detector for complex features cannot always distinguish between one- and two-dimensional stimuli, and simple non-linear post-processing operations such as thresholding or rectification are insufficient to avoid false-positive responses to one-dimensional stimuli. The nonlinearities required to overcome response ambiguities involve operations such as the logical "and". In machine vision, aforementioned has been realized in the logical/linear operators of Iverson and Zucker [3]. Due to their design, local/linear operators respond in a linear fashion when the expected image structure, i.e., an edge or line, is present. Thus a logical-linear operator designed for contour detection has contrast invariance of orientation tuning only if the stimulus is one-dimensional.

However, contrast invariant orientation tuning is an essential ingredient in virtually all population coding models, including that of Zemel and Pillow [8] and the one presented in this paper. An ensemble of non-linear operators would clearly violate the superposition principle and it is expected to exhibit complicated ensemble activity profiles due to the contrast-dependent orientation tuning of its individual units in the presence of multiple orientations. Thus if the logical/linear operators are a more accurate model of orientation selective neurons in the visual cortex than the rectified linear one, a realistic population coding model of the cortical hypercolumn would require more complicated processing strategies incorporating the variability of neural response properties. Clearly, more experiments involving complex stimuli will be necessary to answer these questions.

References

[1] M. Carandini and D.L. Ringach. Predictions of a recurrent model of orientation selectivity. *Vision Research*, 37(21):3061–3071, 1997.

[2] A.P. Georgopoulos, A.B. Schwarz, and R.E. Kettner. Neural population coding of movement direction. *Science*, 233:1416–1419, 1986.

[3] L.A. Iverson and S.W. Zucker. Logical/linear operators for image curves. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 17(10):982–996, October 1995.

[4] K.V. Mardia. *Statistics of Directional Data*. Academic Press, London and New York, 1972.

[5] R. Vogels. Population coding of stimulus orientation by striate cortical cells. *Biological Cybernetics*, 64:25–31, 1990.

[6] R. Zemel and P. Dayan. Distributional population codes and multiple motion models. In *Advances in Neural Information Processing*, volume 11, pages 174–180, 1999.

[7] R. Zemel, P. Dayan, and A. Pouget. Probabilistic interpretation of population codes. *Neural Computation*, 10(2):403–430, 1998.

[8] R. Zemel and J. Pillow. Encoding multiple orientations in a recurrent network. *Neurocomputing*, 32-33:609–616, 2000.

[9] R. Zemel, C.K.I. Williams, and M.C. Mozer. Lending direction to neural networks. *Neural Networks*, 8(4):503–512, 1995.

[10] C. Zetzsche and E. Barth. Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research*, 30(7):1111–1117, 1990.

Figure 1: **Left:** Comparison of certainty measures. The resultant length used in [9] (dashed), “saturates” very soon, whereas ours (solid), based on relative entropy (eqn.11), is more suitable to discriminate between certainties in the range $10 < \kappa < 40$ relevant for real images. **Right:** A real image with 256×256 Pixels and several points at edges, corners and T-junctions.

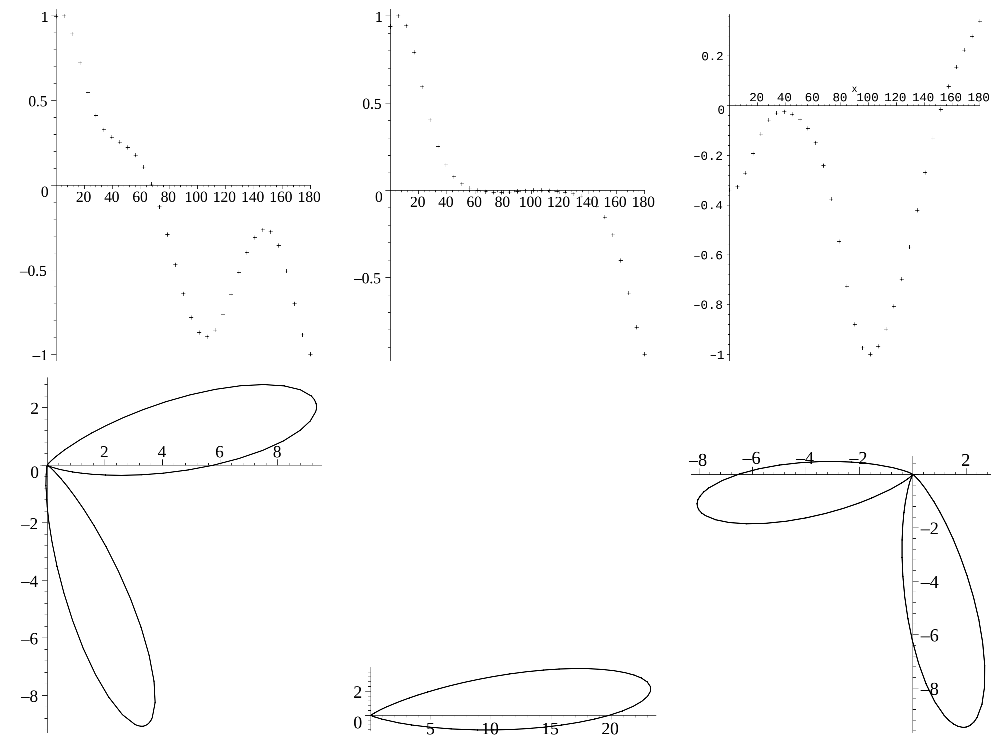

Figure 2: **Top**: Filter bank response profiles at point 1 (left) point 2 (center) and point 5 (right). **Bottom**: Polar plots of the corresponding probability densities.