

A Mixture Model For Population codes of Gabor Filters

Niklas Lütke, Richard C. Wilson
 Department of Computer Science
 University of York
 York, Y010 5DD, UK
 wilson@cs.york.ac.uk

Abstract—Population coding is a coding scheme which is ubiquitous in neural systems, and is also of more general use in coding stimuli, for example in vision problems. A population of responses to a stimulus can be used to represent not only the value of some variable in the environment, but a full probability distribution for that variable. The information is held in a distributed and encoded form which may in some situations be more robust to noise and failures than conventional representations.

Gabor filters are a popular choice for detecting edges in the visual field for several reasons. They are easily tuned for a variety of edge widths and orientations, and are considered a close model of the edge filters in the human visual system. In this paper we consider population codes of Gabor filters with different orientations. A probabilistic model of Gabor filter responses is presented. Based on the analytically derived orientation tuning function and a parametric mixture model of the filter responses in the presence of local edge structure with single or multiple orientations a probability density function of the local orientation in any point (x, y) can be extracted through a parameter estimation procedure.

The resulting pdf of the local contour orientation captures not only angular information at edges, corners or T-junctions but also describes the certainty of the measurement which can be characterized in terms of the entropy of the individual mixture components.

Keywords: Population codes, mixture models, edge detection

I. INTRODUCTION

Population coding seems to be a neural coding scheme of general importance, as it is ubiquitous in the brain [1]. In the population code, the properties of a stimulus are represented by the activities of an ensemble of neurons. Information is distributed over the ensemble rather than being characterised by a single specialised neuron (a so-called "grandmother cell"). An ensemble-based coding scheme has a number of benefits, at the expense of requiring a substantial number of neurons to represent the encoding. The code can approximately represent the full probability distribution of a variable in the environment [2], [3]. In addition, the distributed nature of the representation renders it robust to noise and failure of individual elements.

Probabilistic decoding of sensory input data is concerned with the extraction of a probabilistic information from the responses of a set of filters, each of which is tuned to a particular image feature, for example an edge of particular

orientation[4], [5], [2]. The set of responses as a whole encodes more than just the most likely orientation[6], and may in fact be used to reconstruct a probability density function for the feature in question. The advantage of the probabilistic approach is two-fold. Firstly, it is not restricted to the decoding of a unique value. Secondly, the density function can represent the uncertainty of values. Assuming a mixture distribution for the input variable allows us to represent both uncertainty and ambiguity. Here substantial contributions have been made by Zemel, Dayan and colleagues [2], [7], [8], which are mainly concerned with modelling biological information processing, particularly the perception of motion and orientation.

Sanger has proposed a MLE method[9] which is used to obtain a single, most likely value of the input variable. However, it actually provides an entire conditional distribution $p(x|\mathbf{r}) = \mathcal{L}(x|\mathbf{r})$ in the encoded variable x , given the "vector" of responses \mathbf{r} . However, Zemel et. al showed that the "standard Poisson model" in [9] is incapable of representing densities broader than the tuning curve. Moreover, $p(x|\mathbf{r})$ is always unimodal which makes it impossible to represent ambiguity in x [2].

These serious limitations in representational capacity are overcome in a recently developed more refined probabilistic model by the same authors, which is able to capture multimodal distributions of arbitrary width. Their distributional population coding approach, the "extended Poisson model", is a non-parametric method that aims to recover the encoded probability density as set of sampling points. These discrete values of the probability density are treated as stochastic variables themselves and are determined by maximizing the data likelihood function, thus performing an approximated form of maximum a posteriori estimation in *distributions over distributions*. However, in order to incorporate arbitrary probability densities, both the pdf and the tuning function are approximated by piecewise constant histograms.

The question of complex distributions is very relevant for machine vision where a situation arises when, in addition to simple edges, corner points and junctions are to be detected. Here, classical gradient based edge detection schemes are not efficient, since intensity gradient is ill-defined, and, even when corners are somewhat rounded, detector responses tend to be minute. The usual strategy is to employ special corner detectors, designed to detect two-dimensional intensity features. Zetsche and Barth identified fundamental limitations of

linear filters operating on two-dimensional intensity features [10]. Typically, a linear filter designed to detect 2D-features involves a combination of responses from linear components. They argue that, for signal-theoretic reasons, any such filters will inevitably show false-positive responses to certain one-dimensional stimuli. The response ambiguity cannot be overcome by successive non-linear operations, such as thresholding or rectification. In order to avoid false-positive responses, intrinsic non-linearities, in the form of logical *and*-operations, are necessary, by means of which the linear components are combined.

We argue that a filter bank is capable of representing complex local intensity structure, except for curvature. No additional corner detector is required. In this paper probabilistic decoding techniques are applied to a computer vision task, namely the detection and representation of local edge orientation in computer vision. The ability to represent ambiguous inputs is used to extract multiple orientations in corner points, junctions etc. This contrasts with earlier work, where a population vector method was used to extract tangent fields from unimodal response profiles [11].

Classical edge detectors such as the well known Canny detector [12] usually fail at corner points and T-junctions [12] because the filter responses no longer describe a single edge of definite orientation. Traditional vector-based models of edge orientation and strength are not sufficient at these points because they consist of multiple possible edge orientations.

As opposed to biological models of population coding in which neural firing rates are considered as random variables following a Poisson statistics [2] or a related probabilistic model [13], the filters are treated as deterministic operators fed with a stochastic visual input. Even though the filtering process is deterministic per se the responses themselves become stochastic since they are functions of a random variable, the local contour orientation θ . Because of flexibility of the probabilistic description, the effect of stochastic filters can, to some extent, be accommodated in terms of measured uncertainty in the final result.

In order to recover the whole probability density function(PDF) of the local orientation, $p(\theta)$, which represents the uncertainty and ambiguity of θ , the PDF parameters have to be estimated from the given set of filter responses. This can be achieved indirectly via the PDF of the responses. Based on an empirically and theoretically motivated model of the tuning function of Gabor filters and the assumption of a von Mises mixture distribution of the angular input variable (local contour orientation) the corresponding PDF of the *responses* can be derived and the mixture parameters determined such that the given filter responses are most likely. Thus, the PDF of the input variable is decoded from the activities of the filter outputs.

II. GABOR FILTER RESPONSE PROFILES

Gabor filters are a popular choice for edge detection applications. They can be tuned to respond to particular orientations and feature scales. Furthermore, they are thought to closely resemble the edge filters in the human visual system. They

have the general form

$$\mathcal{G}(x, y) = \exp\left(-[x^2/\sigma_x^2 + y^2/\sigma_y^2]\right) \exp(-2\pi i[ux + vy]) \quad (1)$$

The quantities u and v are related to the frequency and orientation of the image structure under examination. For maximal response they need to be tuned to the dominant frequency and orientation of the edge. The real and imaginary parts are symmetric and anti-symmetric respectively and correspond to line and edge structure. We therefore confine ourselves to the imaginary component since we are concerned with edges. The *sigma* values are dependent on the noise levels in the image.

A. Tuning functions

The tuning function or tuning curve refers to the particular response profile of one member of the population when exposed to the whole range of input stimuli. In this case, that is the response of a Gabor filter when presented with edges of varying orientation. Typically a tuning curve consisted of a single peak around the preferred orientation of the filter. The width of this peak is critical for the performance of the population code, and this aspect has been given much attention in the literature. For example, Zhang and Sejnowski[14] show that for encoding multidimensional stimuli($D \geq 3$), wide tuning curves are favoured because many centres are then activated. For $D = 1$ as discussed in this paper, narrow tuning curves result in more accurate information at each centre, and are therefore favourable. However, Eurich and Wilke[15] demonstrate that tuning widths which are narrow compared to the distance between centres results in regions of stimulus which cannot be accurately represented because they fall between these centres. There is a optimal width for $D = 1$ where there is some degree of overlap between the tuning curves.

Steerable filters[16] provide an efficient way of constructing a large range of oriented filter responses from a small number of basis filters. The idea here is that for certain classes of filter ('steerable filters'), the response at any angle can be calculated from the responses of the basis filters. In this way a filter bank can be constructed from just a few applications of the basis filters to the image. The number of basis filters required is related to the angular bandwidth of the individual filters; if the angular bandwidth is low, few filters are required but limited angular information is recovered[17]. This issue is closely related to the tuning width considerations discussed in the previous paragraph. We require narrow tuning curves with high angular bandwidth to provide accurate orientation information. Gabor filters are useful because they provide a straightforward way to control the width of the tuning curves. Although they are not steerable, provided we choose the size of the filter bank in accordance with the criterion discussed above, we require the same number of filters as would be required to extract the same amount of angular information with a steerable scheme. As an example, Figure (1) shows a image with three orientations along with the angular response curves for a Gabor filter and for the steerable filter derived from the fifth differential of a Gaussian. While the structure is clear in the Gabor filter responses, the steerable filter only finds

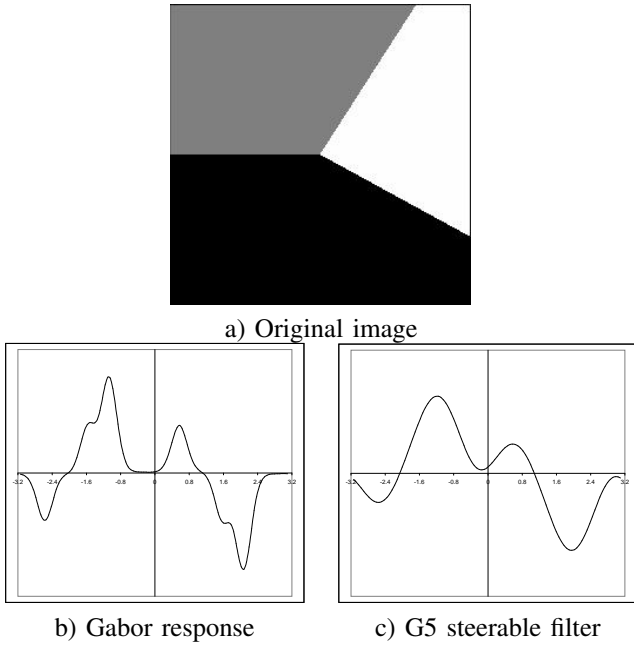


Fig. 1. A multi-orientation image region and the angular responses of the Gabor filter and the fifth differential Gaussian steerable filter

two orientations. To obtain the required angular resolution, we would need more than six basis filters even in a steerable scheme.

The responses of a bank of Gabor filters extracted at locations of ambiguous orientation show multiple peaks that are more or less separated depending on the tuning width of the filters applied. An adequate choice of filter parameters allows to discriminate at least two different response maxima corresponding to the different orientations of the edges meeting or overlapping at the considered point.

B. The tuning function of odd-symmetric Gabor filters

We commence with a model of an edge which is anti-symmetric in one direction about the point of interest and constant in the perpendicular direction. As a consequence, we can expand the image in terms of a Fourier sine series across the edge, i.e.

$$I(x, y) = \sum_i a_i \sin[k_i(x \cos \theta + y \sin \theta)] \quad (2)$$

The response of the filter to such a stimulus is confined to the imaginary part and therefore is linear. As a result, we need only consider the response of an odd symmetric Gabor filter to a sinusoidal grating of arbitrary orientation.

Real images are likely to contain edges that are subject to some degree of blur and therefore have a dominant spatial ground frequency in their spectrum, while higher frequency components are comparatively weak. The windowing effect of the Gaussian in the Gabor filter truncates the low frequency part of the edge structure and so the effective ground frequency is largely dominated by the width of the Gaussian and only impacted to a small extent by the frequency structure of the edge.

For curved stimuli this still holds approximately within the effective range of the filter mask (i.e. where the Gaussian envelope is significantly above value zero).

We consider a sinusoid of orientation θ

$$S(x, y) = \sin[k_s(x \cos \theta + y \sin \theta)]; \quad k_s = 2\pi/\lambda_s. \quad (3)$$

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

$$\mathcal{G}_{odd}(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \sin(k_f x) \quad (4)$$

Clearly, the result for a filter of a preferred orientation ψ may be obtained by rotating the stimulus by $-\psi$. The orientation tuning function $f(\theta)$ is the spatial convolution of the filter with the sinusoid at the origin $(0, 0)$.

$$\begin{aligned} f(\theta) &= (\mathcal{G}_{odd} * S)(0, 0) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{G}_{odd}(-x, -y) S(x, y, \theta) dx dy \\ &\Rightarrow f(\theta) \propto \sinh\left(\underbrace{k_f k_s \sigma^2}_{\kappa_0} \cos \theta\right) \end{aligned} \quad (5)$$

The constant k_s is related to the sharpness of edge transitions in the image, and is assumed to be fixed within any particular image. The tuning curve is therefore the same at any location in the image. The parameter κ_0 is a so-called ‘concentration’ parameter. Its reciprocal value is a measure of the angular variance and therefore controls the orientation tuning width. It depends on the two known filter properties $k_f = 2\pi/\lambda_f$ and σ as well as the unknown quantity $k_s = 2\pi/\lambda_s$ which is inverse proportional to the dominant wavelength λ_s in the input signal. In order to specify the tuning curve, we must select the parameter κ_0 with regard to the sharpness of edge transitions in the image.

The population code consists of a bank of such identical Gabor filters excepting that they have different preferred orientations. If the preferred orientation of filter i in the bank is ψ_i , then the tuning function for that filter is given by

$$f_i(\theta) = C \sinh[\kappa_0 \cos(\theta - \psi_i)] \quad (6)$$

Here $C = 1/\sinh(\kappa_0)$ is a normalization constant, so that $f_i(\theta) \in [-1, 1]$. Figure II-B shows a typical filter response profile for an edge of orientation $\theta = 3$.

III. THEORY OF PROBABILISTIC CODING WITH GABOR FILTERS

The Gabor filter bank is an ensemble of orientation sensitive units. If the filters are strictly linear one can assume that 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

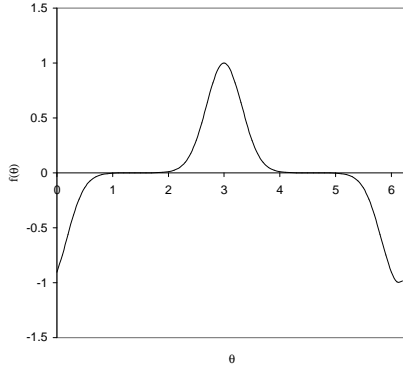


Fig. 2. Typical tuning curve for an edge with $\theta = 3$

activities for the individual edge components. This important assumption is crucial for the following derivations. It is certainly valid for odd or even symmetric Gabor filters but already breaks down with Gabor moduli, since the magnitude of the complex filter response is calculated by a Pythagorean sum, which is a non-linear operation.

A. Model of Line Orientation

While a straight line or edge in the image has a single well defined orientation, problems occur when either the line is curving, or multiple orientations exist within the field of the filter. In addition, noise in the image and edge blurring can create uncertainties even in the angle of a straight line. We can accommodate all these possibilities by adopting a stochastic model of line orientation where the line angle θ is governed by probability density function $p(\theta)$. Such a model can incorporate noise and multiple orientations, and also to some extent curvature in the sense that curvature is signalled as a range of possible angles. Although this model does not explicitly incorporate stochastic filter responses, such as those present in neural population codes, such uncertainty may be implicitly accommodated by increased uncertainty in the resulting angular PDF.

B. Expectation values of filter responses

Let ψ be the preferred orientation of a filter. Assuming that superposition holds, 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 $p(\theta)$ [2]:

$$\bar{r}(\psi) = \int_0^{2\pi} f(\psi - \theta)p(\theta) d\theta. \quad (7)$$

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 ensemble activities (the response profile) are a sample of this function for a discrete set of filter orientations ψ_ν . In the following sections we will derive a parametric model of the expected filter responses based on the orientation tuning function and a mixture model of $p(\theta)$.

C. A mixture model of local orientation

The encoding equation (7) is general in the sense that it does not restrict $p(\theta)$ in any way. We now choose a particular form for the PDF of line 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[18]. The von Mises distribution, which is essentially an orientation based analogue of the Gaussian distribution, models each principal line orientation with a mean value and a variance. By choosing this model, we are assuming that a) the image consists of edges of well defined orientation, and b) uncertainties in these orientations can be modelled without higher order statistics. By using a mixture of these distributions, multiple principle orientations can be modelled. Equation 8 gives the appropriate formula for this model.

$$p(\theta) = \frac{1}{2\pi} \sum_{i=1}^m \frac{P(i)}{I_0(\kappa_i)} e^{\kappa_i \cos(\theta - \bar{\theta}_i)}, \quad \text{with } \sum_{i=1}^m P(i) = 1 \quad (8)$$

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. Equation (8) can be considered as being an circular analogon of the Gaussian mixture density. The κ_i are analogous to the inverse variance and the $\bar{\theta}_i$ to the mean in a Gaussian PDF. The $P(i)$ are the mixing coefficients.

Edge structure is quite constrained in the sense that there are a limited number of orientations which can exist at a point in an image. A straight edge will clearly have a single orientation. Where this edge meets another on the boundary of an object, two orientations will exist at the corner point. If the edge lies over another edge in a background object (occluding edge), again two orientations will exist. Three orientations will only exist at the vertex of a cuboid object or where a corner aligned exactly with an occluded edge. The number m of mixture components will be limited to two which allows to describe the essential cases of multiple edge orientation, i.e. corners and T-junctions.

Inserting equations (6) and (8) in (7) yields the expectation value of the response profile as a function of the tuning width and the PDF parameters (the hyperbolic sine is written in exponential form):

$$\bar{r}(\psi) = \frac{C}{4\pi} \sum_{i=1}^m \left(\frac{P(i)}{I_0(\kappa_i)} \int_0^{2\pi} \left[e^{\kappa_0 \cos(\theta - \psi)} - e^{-\kappa_0 \cos(\theta - \psi)} \right] e^{\kappa_i \cos(\theta - \bar{\theta}_i)} d\theta \right)$$

Integration yields:

$$\bar{r}(\psi) = \frac{C}{2} \sum_{i=1}^m \frac{P(i)}{I_0(\kappa_i)} \left\{ I_0 \left(\sqrt{\kappa_0^2 + \kappa_i^2 + 2\kappa_0\kappa_i \cos(\psi - \bar{\theta}_i)} \right) - I_0 \left(\sqrt{\kappa_0^2 + \kappa_i^2 - 2\kappa_0\kappa_i \cos(\psi - \bar{\theta}_i)} \right) \right\} \quad (9)$$

This then is the expectation value of the responses of our Gabor filter bank under the mixture of von Mises distribution model. However, because of the uncertainties created by image noise, curving edges and edge blur, the actual responses of the Gabor filter bank will deviate from the expectation value. In order to recover the maximum likelihood model of orientation we need to know these uncertainties via the probability distribution of responses.

IV. THE PROBABILITY DISTRIBUTION OF RESPONSES

In order to implement a maximum likelihood estimation scheme for the parameters of the mixture model, we need to evaluate probabilities of the filter responses rather than just the expectation value. In other words, it is essential to derive the response PDF of the individual filters, i.e. $p_\nu(r_\nu)$. The filter responses r_ν obtained at a certain location in an image are samples of the random variable $r(\psi)$ at a particular instant for the different orientations in the filter bank ψ_ν , $\nu = 1 \dots n$. The density $p_\nu(r_\nu)$ depends on the parameters of the mixture PDF $p(\theta)$ plus the parameters specifying the tuning function, in particular λ_s , the dominant wavelength in the input signal. The mixture parameters can then be found through a maximum likelihood estimation using standard techniques such as the EM algorithm. In principle there are two ways of finding the parameters of the tuning function:

- a) from a previous measurement using test stimuli (straight lines)
- b) together with the mixture parameters in the same MLE process

While b) is an interesting approach, for the purposes of this paper, λ_s , and hence k_s , is determined from straight line portions of the test images.

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 (6). According to the general theorem for the probability density of a function of a random variable (see e.g. [19]) the PDF of the response profile (preferred orientation ψ) is given by:

$$p_\nu(r) = \frac{p(\theta_1)}{|f'_\nu(\theta_1)|} + \frac{p(\theta_2)}{|f'_\nu(\theta_2)|}, \quad (10)$$

where f' is the derivative of the tuning function with respect to θ , and $\theta_{1,2}$ are the two corresponding angles of the response value r in the interval $[\psi_\nu - \frac{\pi}{2}, \psi_\nu + \frac{\pi}{2}]$ given by the inverse of the tuning function, i.e. the solutions of the equation:

$$r = f_\nu(\theta_{1,2}), \quad \theta_{1,2} \in [\psi_\nu - \frac{\pi}{2}, \psi_\nu + \frac{\pi}{2}] \quad (11)$$

The resulting distribution turns out to be

$$p_\nu(r) = \frac{1}{\pi C \kappa_0} \sum_{i=1}^m \frac{P(i)}{I_0(\kappa_i)} \frac{\exp \left[\frac{\kappa_i}{\kappa_0} \sinh^{-1} \left(\frac{r}{C} \right) \cos(\psi - \bar{\theta}_i) \right]}{\sqrt{1 + \left(\frac{r}{C} \right)^2} \sqrt{1 - \left[\frac{1}{\kappa_0} \sinh^{-1} \left(\frac{r}{C} \right) \right]^2}} \cdot \cosh \left(\kappa_i \sqrt{1 - \left[\frac{1}{\kappa_0} \sinh^{-1} \left(\frac{r}{C} \right) \right]^2} \sin(\psi - \bar{\theta}_i) \right) \quad (12)$$

This density is again a mixture model. However, the mixture components are highly non-Gaussian which makes it more difficult, though not impossible, to find the parameters.

V. PARAMETER ESTIMATION VIA THE EM-ALGORITHM

Having derived the PDF of responses from the PDF of the local orientation, it is possible to estimate the model parameters with the EM-algorithm. However, due to the complexity of the response PDF the log-likelihood does not lead to significant simplification as in Gaussian case. Instead, some update equations are transcendental and require a numerical procedure in order to solve for the new parameter values.

Adopting to the notation used in [20] the quantity to minimize is:

$$\tilde{Q} = - \sum_{\nu=1}^n \sum_{i=1}^m P^{old}(i|r_\nu) \ln \{ P^{new}(i) p^{new}(r_\nu|i) \} \quad (13)$$

Here $p^{new}(r_\nu|i)$ is the i -th mixture component of (12) for $\psi = \psi_\nu$. According to Bayes theorem the ‘‘old’’ posterior probabilities $P^{old}(i|r_\nu)$, which describe the label assignment of data points to individual mixture components, are given by[20]:

$$P^{old}(i|r_\nu) = \frac{p^{old}(r_\nu|i) P^{old}(i)}{p^{old}(r_\nu)} \quad (14)$$

where $p^{old}(r_\nu) = p^{old}(r; \psi_\nu)$ is given by (12). This is the expectation step. In the maximization step, the update equations for the mixture parameters are obtained by differentiating (13) with respect to these parameters and setting the derivative to zero.

In the case of the mixing coefficients $P^{new}(i)$ the update equation has to be derived under the constraint $\sum_{i=1}^m P^{new}(i) = 1$ by means of a Lagrange multiplier[20], which leads to:

$$P^{new}(i) = \frac{1}{n} \sum_{i=1}^n P^{old}(i|r_\nu) \quad (15)$$

Let Θ_i be the remaining mixture parameters. In order to obtain their update equations it is necessary to take the derivative of \tilde{Q} with respect to the Θ_i^{new} :

$$\frac{\partial \tilde{Q}}{\partial \Theta_i^{new}} = - \sum_{\nu=1}^n \sum_{k=1}^m P^{old}(k|r_\nu) \frac{\partial}{\partial \Theta_i^{new}} \ln p^{new}(r_\nu|k) \quad (16)$$

However, the derivative with respect to Θ_i^{new} is only non-vanishing if $k = i$. Thus, the general form of the update equation for Θ_i becomes:

$$-\sum_{\nu=1}^n P^{old}(i|r_\nu) \frac{\partial}{\partial \Theta_i^{new}} \ln p^{new}(r_\nu|i) = 0 \quad (17)$$

According to (12), and using the abbreviation $\xi_\nu = \frac{1}{\kappa_0} \sinh^{-1} \left(\frac{r_\nu - B}{C} \right)$ for the frequently occurring ‘‘data term’’, the logarithm of the i -th mixture component for response r_ν is:

$$\begin{aligned} \ln p^{new}(r_\nu|i) = & -\ln \{ \pi C \kappa_0 I_0(\kappa_i^{new}) \} - \frac{1}{2} \ln \{ 1 - \xi_\nu^2 \} \\ & + \kappa_i^{new} \xi_\nu \cos(\psi_\nu - \bar{\theta}_i^{new}) - \frac{1}{2} \ln \left\{ 1 + \left(\frac{r_\nu}{C} \right)^2 \right\} \\ & + \ln \cosh \left(\kappa_i^{new} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right) \end{aligned} \quad (18)$$

Inserting (18) into (17) yields the derivative of \tilde{Q} with respect to the angular modes $\bar{\theta}_i$:

$$\begin{aligned} \frac{\partial \tilde{Q}}{\partial \bar{\theta}_i^{new}} = & \sum_{\nu=1}^n P^{old}(i|r_\nu) \left\{ \kappa_i^{new} \xi_\nu \sin(\psi_\nu - \bar{\theta}_i^{new}) \right. \\ & \left. - \kappa_i^{new} \sqrt{1 - \xi_\nu^2} \cos(\psi_\nu - \bar{\theta}_i^{new}) \cdot \right. \\ & \left. \tanh \left(\kappa_i^{new} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right) \right\} \end{aligned}$$

Dividing by κ_i^{new} and thereby excluding the irrelevant solution $\kappa_i^{new} = 0$ (i.e. uniform angular distribution, no directional structure) yields a transcendental update equation in $\bar{\theta}_i^{new}$ and κ_i^{new} :

$$\begin{aligned} \sum_{\nu=1}^n P^{old}(i|r_\nu) \left\{ \xi_\nu \sin(\psi_\nu - \bar{\theta}_i^{new}) \right. \\ \left. - \sqrt{1 - \xi_\nu^2} \cos(\psi_\nu - \bar{\theta}_i^{new}) \cdot \right. \\ \left. \tanh \left(\kappa_i^{old} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right) \right\} = 0 \end{aligned} \quad (19)$$

This equation is the first of a pair of transcendental equations governing the parameter update. Using $\frac{d}{dx} I_0(x) = I_1(x)$ the derivative of (18) with respect to κ_i^{new} is:

$$\begin{aligned} \frac{\partial}{\partial \kappa_i^{new}} \ln p^{new}(r_\nu|i) = & -\frac{I_1(\kappa_i^{new})}{I_0(\kappa_i^{new})} + \xi_\nu \cos(\psi_\nu - \bar{\theta}_i^{new}) \\ & + \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \\ & \tanh \left(\kappa_i^{new} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right) \end{aligned}$$

Hence the second update equation is:

$$\begin{aligned} \sum_{\nu=1}^n P^{old}(i|r_\nu) \left\{ \frac{I_1(\kappa_i^{new})}{I_0(\kappa_i^{new})} - \xi_\nu \cos(\psi_\nu - \bar{\theta}_i^{new}) \right. \\ \left. - \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right. \\ \left. \tanh \left(\kappa_i^{new} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right) \right\} = 0 \end{aligned} \quad (20)$$

Since κ_i^{new} appears as the argument of modified Bessel functions and the hyperbolic tangent the update equation is again transcendental. Thus, for each mixture component (19) and (20) form a system of transcendental update equations which have to be solved numerically for the pair $(\bar{\theta}_i^{new}, \kappa_i^{new})$ of new parameter values within each maximization step, e.g. using a Newton-Raphson algorithm. While these equations may appear daunting, they are easily solved numerically since they converge quickly to the correct solution. It is this approach which is used in the experimental section. However, a closed form approximation can be developed and this is examined in the next section.

A. An approximation for closed form update equations

The transcendental nature of the update equations (19) and (20) stems from the fact that parameters appear inside a hyperbolic tangent and a ratio of Bessel functions, respectively. Therefore it would be desirable to substitute these expressions by suitable approximations. For practical purposes the ratio of modified Bessel functions $I_1(\kappa_i^{new})/I_0(\kappa_i^{new})$ can be replaced by its asymptotic approximation $1 - \frac{1}{2\kappa}$. The approximation is rather benign with accuracy better than 1% for $\kappa > 4.5$ and turns out to be very suitable for the range of κ values found in real images.

The terms with the hyperbolic tangent in equation (19) can be simplified by replacing the new parameter values by their values from the previous update step, i.e.

$$\begin{aligned} \tanh \left(\kappa_i^{new} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right) \\ \approx \underbrace{\tanh \left(\kappa_i^{old} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{old}) \right)}_{D_\nu} \end{aligned}$$

Using the abbreviation D_ν for these now independent terms the update equation for $\bar{\theta}_i^{new}$ becomes

$$\begin{aligned} \tan(\bar{\theta}_i^{new}) = & \frac{\sum_{\nu=1}^n P^{old}(i|r_\nu) \left\{ \xi_\nu \sin \psi_\nu - D_\nu \sqrt{1 - \xi_\nu^2} \cos \psi_\nu \right\}}{\sum_{\nu=1}^n P^{old}(i|r_\nu) \left\{ \xi_\nu \cos \psi_\nu + D_\nu \sqrt{1 - \xi_\nu^2} \sin \psi_\nu \right\}} \end{aligned} \quad (21)$$

The simplification of the update equation for the κ_i^{new} is achieved in a similar way, only that now $\bar{\theta}_i^{new}$ is known and can be used to define

$$\tilde{D}_\nu := \tanh \left(\kappa_i^{old} \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \right).$$

With the asymptotic approximation, equation (20) becomes

$$\kappa_i^{new} = \frac{1}{2} \sum_{\nu=1}^n P^{old}(i|r_\nu) \left/ \sum_{\nu=1}^n P^{old}(i|r_\nu) \left\{ 1 - \xi_\nu \cos(\psi_\nu - \bar{\theta}_i^{new}) - \sqrt{1 - \xi_\nu^2} \sin(\psi_\nu - \bar{\theta}_i^{new}) \tilde{D}_\nu \right\} \right. \quad (22)$$

The approximation sets a certain lower bound for the value of κ_i meaning that the angular width of the mixture components must not be too broad. This is, however, in agreement with the requirements for separability of modes in the mixture distribution. In other words, the approximation is most accurate when the multiple values for local orientation can be clearly distinguished, i.e. when the certainty of the multiple orientation measurement is sufficiently high.

VI. MEASURES OF CERTAINTY IN PROBABILISTIC POPULATION CODING

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 quality or certainty of these measurements through the concentration parameters of the mixture, the κ_i . However, the κ_i 's cannot directly be interpreted as a measure of the certainty of an orientation estimate. A certainty measure γ should be a function of the variance of the angular estimate. Also, it is desirable for the certainty measure to be normalized and positive, i.e $0 \leq \gamma \leq 1$.

Unlike in the case of a Gaussian pdf where the parameter σ^2 is the variance, the corresponding values $1/\kappa_1, \dots, 1/\kappa_m$ in the von Mises mixture, however, cannot be interpreted as angular variances themselves.

In fact, the concepts of mean and variance cannot be uniquely extended to circular statistics but have to be redefined with some care. For an angular variable the *circular variance* V_0 is defined as (see e.g. [18])

$$V_0 = 1 - E\{\cos(\theta - \bar{\theta})\}. \quad (23)$$

where $E\{\cdot\}$ denotes the expectation value and $\bar{\theta}$ the mean direction. The circular variance is bounded, and $0 \leq V_0 \leq 1$ since the probability mass is concentrated in the interval $[\bar{\theta} - \frac{\pi}{2}, \bar{\theta} + \frac{\pi}{2}]$. Furthermore, V_0 is invariant with respect to a shift of the mean direction, as one would expect from a variance. A convenient choice for a certainty measure is

$$\rho = 1 - V_0 = E\{\cos(\theta - \bar{\theta})\}, \quad (24)$$

which is also referred to as the *resultant length* [18]. ρ inherits the property of shift invariance with respect to the mean direction and also fulfills $0 \leq \rho \leq 1$. For a von Mises distribution this yields $\rho = \frac{I_1(\kappa)}{I_0(\kappa)}$. This measure has been proposed by Zemel and colleagues within the framework of the ‘‘directional-unit Boltzmann machine’’ [21]. However, for the values of κ_i which result from the edges in images, this

measure is usually very close to 1 and does not provide much discriminating power. For this reason, we investigate an alternative.

A. A normalized certainty measure based on entropy

In order to derive a positive certainty measure from entropy it is useful to consider the Kullback-Leibler divergence of a component density $p_i(\theta)$ from the distribution of maximum entropy, i.e. the uniform distribution $q(\theta) = 1/2\pi$. Since the uniform distribution is, by definition, a constant this is equivalent to taking the difference between the maximum entropy (corresponding to $\kappa = 0$) and the entropy for a given $\kappa = \kappa_i$, which always results in a positive quantity. Therefore, the Kullback-Leibler divergence

$$K(p, q) = \int_0^{2\pi} p(\theta) \ln \left[\frac{p(\theta)}{q(\theta)} \right] d\theta. \quad (25)$$

is also called *relative entropy*. With the densities given by

$$p(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \bar{\theta})} \quad \text{and} \quad q(\theta) = \frac{1}{2\pi}$$

and since $q = const.$ this yields:

$$\begin{aligned} K(p_i, \frac{1}{2\pi}) &= \int_0^{2\pi} p_i(\theta) [\ln p_i(\theta) + \ln(2\pi)] d\theta \\ &= -h(p) + \ln(2\pi) \\ &= h\left(\frac{1}{2\pi}\right) - h(p) \end{aligned}$$

This equates to

$$\begin{aligned} K(p_i, \frac{1}{2\pi}) &= \frac{I_1(\kappa_i)}{I_0(\kappa_i)} \kappa_i - \ln[2\pi I_0(\kappa_i)] + \ln[2\pi] \\ &= \frac{I_1(\kappa_i)}{I_0(\kappa_i)} \kappa_i - \ln[I_0(\kappa_i)] \end{aligned} \quad (26)$$

Normalization can be achieved through a squashing function, such as a variation of the familiar logistic sigmoid function

$$g(x) = \frac{2}{1 + e^{-x}} - 1$$

The resulting certainty measure is

$$\gamma = g[K(p_i, \frac{1}{2\pi})] = \frac{2}{1 + I_0(\kappa) \exp[-\kappa I_1(\kappa)/I_0(\kappa)]} - 1 \quad (27)$$

Fig.(3) shows a plot of the two certainty measures, the squashed Kullback-Leibler divergence γ and the resultant length ρ . Compared to the resultant length this measure is more suitable to discriminate measurements by certainty, as it approaches its asymptote of perfect certainty slowly for a very large concentration parameter. The resultant length rapidly approaches the asymptote, assigning a certainty of $\rho \approx 1$ to κ -values above 10. However, for real images the range $10 < \kappa < 40$ turns out to be very important and the entropy-based certainty measure allows a better judgement of the reliability of orientation measurements.

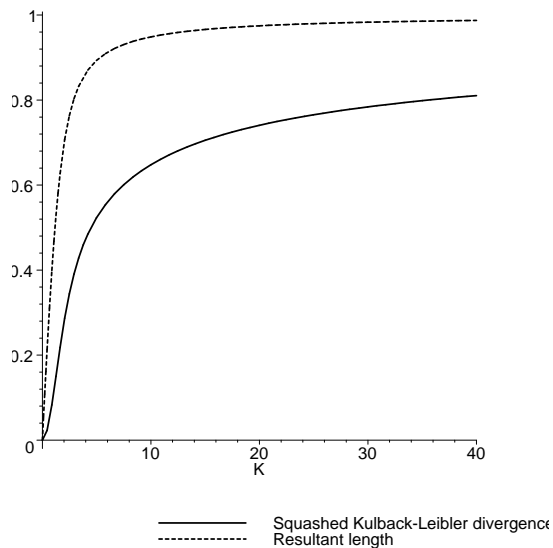


Fig. 3. Comparison of the resultant length $\rho = I_1(\kappa)/I_0(\kappa)$, proposed as a certainty measure by Zemel and colleagues, with the certainty measure γ based on relative entropy (Kullback-Leibler divergence) given by eqn.(27). The resultant length “saturates” too soon whereas γ is more suitable to discriminate between different certainties in the range of $10 < \kappa < 40$, which is important for real images.

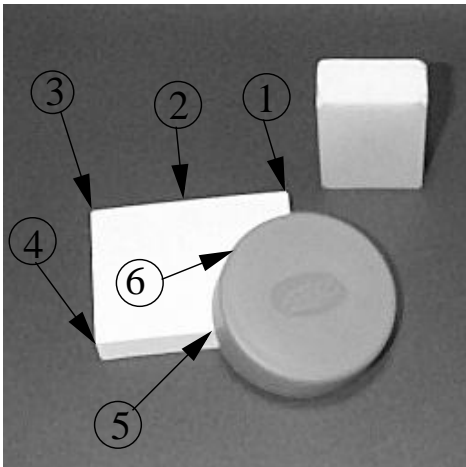


Fig. 4. A real image with 256×256 Pixels and several points at edges, corners and T-junctions.

VII. RESULTS

A. Edge configurations

Figure 4 shows an image containing three objects which generate a number of different edge configurations, including straight edges, curved edges, corners and T-junctions.

We have applied the EM-algorithm described in the previous sections to a number of key points in this image. Figure 5 shows the resulting mixture model from the application of the EM-algorithm. The plot is a polar plot, i.e. the angle represents θ and the radius $p(\theta)$. Point (1) is a corner point and as expected, two narrow perpendicular components are found.

Point (2) is a straight edge. The EM-algorithm finds a strong component in the direction of the edge. Point (3) is again a corner with different orientation. Again two strong components are found. Point (4) shows a strong edge and a

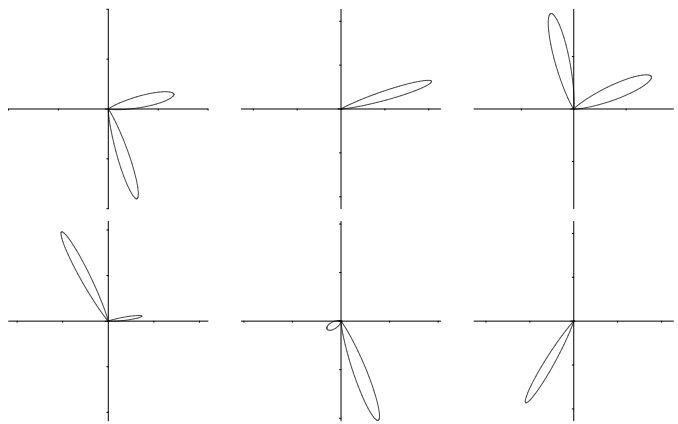


Fig. 5. The solutions for the points labelled 1(top left),2(top centre),3(top right),4,5 and 6 in image 4. The results are plotted in polar form.

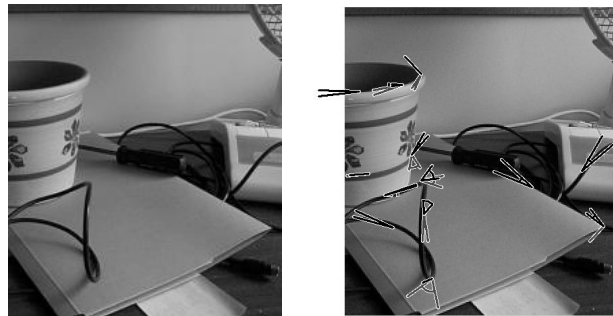


Fig. 6. An office desk with the mixture components at various points marked

weak secondary edge corresponding to the faint internal line of the block. There are in fact three components at point (4) but two of the components are closely aligned and cannot be distinguished by our algorithm. Point (5) 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 in figure 5(right), we get a large component for the curved edge and a smaller component for the weaker secondary edge. Finally, point (6) is a curved edge. The curvature of edges in this image is not sufficient to have much impact on the component width.

The second image is an image of an office desk (figure 6). Several points of interest are marked on the image. These points are marked with the components of the mixture model found at that point. Each component is represented by a pair of lines; the length of the lines represents the strength of the component and the angle between the lines denoted the width of the component. Straight edges are represented by single narrow components. Corner points where the direction is discontinuous are found as 2-component mixtures corresponding to the individual edge directions. In contrast, curves appear either as a wide single component, or beyond a critical curvature, as two components. However, they can be distinguished from corners by the fact that the two components have considerable overlap compared to corner points.

B. Curvature, Noise and Certainty

By employing a model of line orientation based on a probabilistic description of uncertainty, we aim to achieve two

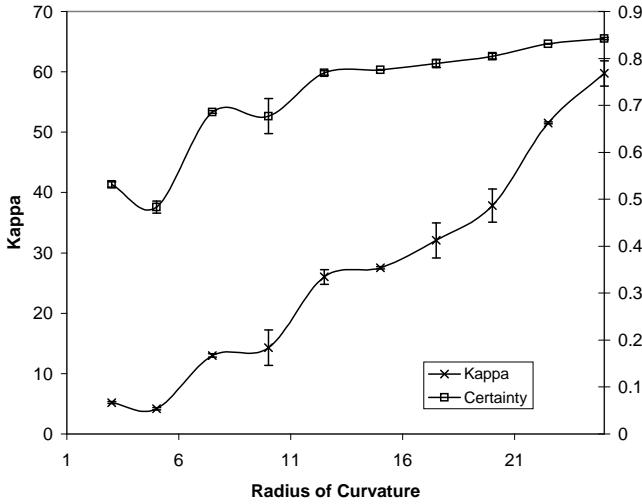


Fig. 7. Width and certainty for a set of circles of varying curvatures

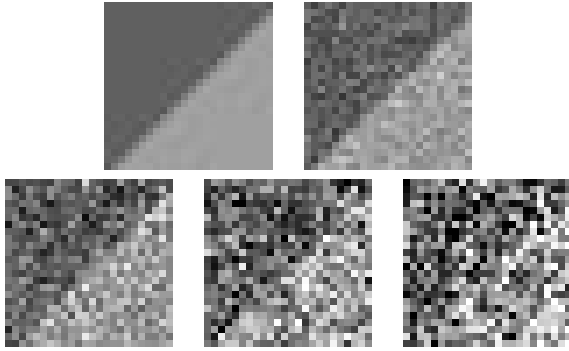


Fig. 8. A short line segment with added Gaussian noise of levels 0,16,32,48 and 64 respectively

objectives. The first is to accommodate various sources of error including image noise and modelling errors (for example we expect straight lines, not curves). The second aim is to correctly assess the degree of uncertainty in the estimate. We now investigate these properties with a number of experiments.

1) *Circles*: The ability of the algorithm to characterise curving lines can be investigated by presenting a number of circles with different radii. We have chosen circles with radii ranging from 25 pixels to 5 pixels. The parameters of the Gabor filter are $\sigma = 6$ and $\lambda_f = 12$, and the tuning function parameter is $\kappa_0 = 12$. For all of these circles, the 45 deg angle on the upper left edge of the circle is correctly estimated. Figure 7 shows a plot of the recovered width parameter κ_1 and the corresponding certainty measure. The distribution width increases steadily with the curvature. The certainty of the measurement remains high however, until the radius of curvature drops below the σ of the Gabor filters.

2) *Image noise*: In order to evaluate the effect of image noise on our method, we use a short line segment. We have added various amounts of Gaussian noise to the image, examples of which are shown in figure 8. The signal level is 64, and noise levels up to a signal-to-noise ratio of 1:1 are investigated.

The aim here is to correctly estimate the orientation of the line and the associated uncertainty. Figure 9 shows the

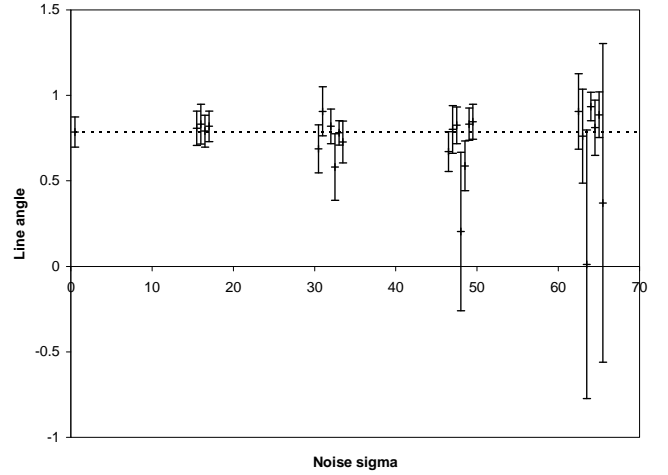


Fig. 9. Estimated line angles and their standard deviations from a noisy line segment



Fig. 10. Low angle corners, with angles θ_c of 0.66, 0.46, 0.27 and 0.15 radians respectively

estimated line angle for noise levels 0,16,32,48 and 64. There are a number of samples for each noise level, and these samples are each offset slightly along the x-axis to make them easy to separate on the plot (for example all samples close to 16 on the plot have a noise level of exactly 16). The error bars on the points represent the standard deviations of the extracted von-Mises distribution. The accuracy of the estimate decreases with increasing noise level. For all but four samples, the true angle lies within one standard deviation of the mean, and is always within two standard deviations.

C. Corner angles

Finally, we test the ability of the algorithm to discriminate low-angle corners. Here we use two components in the mixture model and a larger Gabor filter $\sigma = 12$, $\lambda = 21$ to mitigate the stair-step effects of low-angle lines. Typical stimuli are shown in figure 10. The results are shown in table I. The ideal result in each case is $\theta_1 = 0$, $\theta_2 = \theta_c$, and $P_1 = P_2 = 0.5$. For very low angles, the second component is absorbed into the first which finds the average angle of the corner.

Corner angle θ_c	θ_1	P_1	θ_2	P_2
0.66	-0.02	0.50	0.70	0.50
0.56	-0.02	0.51	0.62	0.49
0.46	0.07	0.66	0.55	0.34
0.37	0.05	0.79	0.48	0.21
0.27	0.02	0.79	0.39	0.21
0.15	0.05	1.0	-	-

TABLE I
RESULTS OF THE MIXTURE MODEL FOR LOW ANGLE CORNERS

VIII. CONCLUSIONS

Starting from a bank of Gabor filter responses at different orientations, we have developed a probabilistic model of these responses. By treating the edge orientation at a point as a stochastic variable governed by a mixture model of von Mises distributions, we are able to model multiple orientations and uncertainty in the edge field. We have then developed an EM algorithm to establish the parameters of the mixture model given a bank of filter responses. We have demonstrated its effectiveness at extracting accurate angle information and multiple orientations on various edge configurations in real images. We have also analysed the impact of curvature of the line segments and image noise on the method and shown that the resulting orientation distributions are consistent with the structure and uncertainties in the images.

REFERENCES

- [1] S. Tanaka and N. Nakayama, "Numerical simulation of neural population coding: Influences of noise and tuning width on the coding error," *Biological Cybernetics*, vol. 73, pp. 447–456, 1995.
- [2] R. S. Zemel, P. Dayan, and A. Pouget, "Probabilistic interpretation of population codes," *Neural Computation*, vol. 10, no. 2, 1998.
- [3] G. E. Hinton, "How neural networks learn from experience," *Scientific American*, vol. 267, no. 3, pp. 145–151, 1992.
- [4] C. H. Anderson, "Basic elements of biological computational systems," *International Journal of Modern Physics C*, vol. 5, no. 2, pp. 135–137, 1994.
- [5] C. H. Anderson and D. C. Van Essen, "Neurobiological computational systems," in *Computational Intelligence Imitating Life*, IEEE Press, pp. 213–222, 1994.
- [6] E. N. Brown, L. M. Frank, D. D. Tang, M. C. Quirk, and M. A. Wilson, "A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells," *Journal of Neuroscience*, vol. 18, no. 18, pp. 7411–7425, 1998.
- [7] R. Zemel and J. Pillow, "Encoding multiple orientations in a recurrent network," *Neurocomputing*, vol. 32–33, pp. 609–616, 2000.
- [8] R. Zemel and P. Dayan, "Distributional population codes and multiple motion models," in *Advances in Neural Information Processing*, vol. 11, 1999, pp. 174–180.
- [9] D. Sanger, "Probability density estimation for the interpretation of neural population codes," *Journal of Neurophysiology*, vol. 76, no. 4, pp. 2790–2793, 1996.
- [10] C. Zetsche and E. Barth, "Fundamental limits of linear filters in the visual processing of two-dimensional signals," *Vision Research*, vol. 30, no. 7, pp. 1111–1117, 1990.
- [11] N. Lüdtkke, R. Wilson, and E. Hancock, "Tangent fields from population coding," in *Proceedings of the IEEE International Workshop on Biologically Motivated Computer Vision (BMCV2000)*, 2000.
- [12] J. Canny, "A computational approach to edge detection," *IEEE Transact. on Patt. Rec. and Machine Intell.*, vol. 8, no. 6, pp. 679–700, 1986.
- [13] L. Itti, J. Braun, D. Lee, and C. Koch, "A model of early visual processing," in *Advances in Neural Information Processing*, vol. 10, 1998, pp. 173–179.
- [14] K. Zhang and T. J. Sejnowski, "Neuronal tuning: To sharpen or broaden?" *Neural Computation*, vol. 11, pp. 75–84, 1999.
- [15] C. Eurich and S. D. Wilke, "Multidimensional encoding strategy of spiking neurons," *Neural Computation*, vol. 12, pp. 1519–1529, 2000.
- [16] W. T. Freeman and E. H. Adelson, "Design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.
- [17] W. C. Yu, K. Daniilidis, and G. Sommer, "Approximate orientation steerability based on angular gaussians," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 193–205, 2001.
- [18] K. Mardia, *Statistics of Directional Data*. London and New York: Academic Press, 1972.
- [19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1984.
- [20] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press, 1995.
- [21] R. Zemel, C. Williams, and M. Mozer, "Lending direction to neural networks," *Neural Networks*, vol. 8, no. 4, pp. 503–512, 1995.
- [22] N. Lüdtkke, R. Wilson, and E. Hancock, "Population codes for orientation estimation," in *Proc. of the International Conference on Pattern Recognition*, 2000.
- [23] R. Wilson and N. Lüdtkke, "Decoding population codes," in *Proc. of the International Conference on Pattern Recognition*, vol. 2, 2000, pp. 137–140.



Niklas Lüdtkke was born in Berlin, Germany in 1969. He graduated in theoretical Physics from the University of Kiel, Germany in 1995. After 13 months of civil service he studied computer vision and pattern recognition at the Institute for Neural Computation in Bochum, Germany, with Christoph von der Malsburg. He then went to England to join the group of Prof. Edwin Hancock at the University of York, where he received his PhD in computer vision in 2003. Currently, he is a postdoctoral research associate in Prof. Mark Nelson's lab at the Beckman

Institute of the University of Illinois, USA.

His research interests include statistical pattern recognition, biologically inspired computer vision and models of neural coding and information processing.



Richard Wilson received the BA degree in Physics from the University of Oxford in 1992. In 1996 he received the DPhil degree from the University of York for his thesis "Inexact graph matching using symbolic constraints". From 1996 to 1998 he worked as a Research Associate at the University of York. After a period of postdoctoral research, he was awarded an Advanced Research Fellowship in 1998, a position which he currently holds in the Department of Computer Science at the University of York.

He has published some 70 papers in journals, edited books and refereed conferences. He received an outstanding paper award in the 1997 Pattern Recognition Society awards and has won the best paper prize in ACCV 2002. He is currently an Associate Editor of the journal Pattern Recognition. His research interests are in statistical and structural pattern recognition, graph methods for computer vision, high-level vision and scene understanding. He is a member of the IEEE computer society.