A Population Coding Approach to Edge Detection and Perceptual Grouping

Niklas Lüdtke

Submitted for the degree of Doctor of Philosophy

Department of Computer Science

The University of York

30/09/2002

Abstract

This dissertation presents a novel, biologically inspired approach to edge detection and perceptual organisation, based on a synthesis of the well-known Gabor filters with the concept of population coding from computational neuroscience.

A Gabor filter bank is regarded as an ensemble of orientation sensitive units that encode local contour orientation in a distributed fashion, somewhat akin to the "simple cells" in the mammalian primary visual cortex. From the filter ensemble, a probability density function (pdf) of local contour orientation is decoded by taking into account the orientation tuning function of the filters and assuming a von Mises mixture model for the contour angle. The parameters of the pdf are estimated using an expectation maximisation (EM) algorithm.

Whereas conventional edge detection schemes tend to reduce the set of filter responses in each pixel to a single quantity, e.g. a local tangent angle, this dissertation takes a different approach, aiming to maintain a distributed representation. The benefits of the resulting analytically derived probabilistic population decoding algorithm is that points with multiple orientations, such as corner points or junctions, can be accommodated within the same framework by means of *multimodal* probability densities.

Another important aspect of distributed coding is the notion of *certainty*, characterised by the spread of activity across the filter bank or the entropy of the orientation pdf. It is demonstrated that the availability of local feature certainty *prior* to perceptual organisation is beneficial for feature localisation. Selecting features by means of their certainty, rather than by thresholding filter responses, renders the feature extraction contrast independent and more robust against noise.

In the subsequent grouping step, small curved contour segments are generated through spline interpolation between pairs of locally extracted tangent elements. The grouping process involves a revision of the local orientation measurements, controlled

by their certainty values and the overall curvature of the connecting spline. This is possible only because certainty has become a *measured* quantity determined at the stage of local feature extraction. In most other grouping schemes, certainty is either not considered, or, as in probabilistic relaxation labeling, inferred *after* feature extraction during the optimisation of probability density parameters representing local features.

While not claiming to present a model of biological visual processing, this thesis provides some new insight into the initial problems that both artificial and neural visual systems are confronted with: the extraction and representation of local features following sensory acquisition, and the subsequent grouping of such locally extracted features into larger, more complex entities.

Contents

Declaration

I hereby declare that, except where indicated, the work presented in this thesis is entirely my own. Some results have been published previously (Lüdtke and Hancock, 2001; Lüdtke et al., 2000a; Lüdtke et al., 2000b; Lüdtke et al., 2002b; Lüdtke et al., 2002a); others have been submitted for publication (see list of publications).

List of Publications

Wilson, R.C. and Lüdtke, N. (2000). Decoding Population Codes. In *Proceedings of the 15th International Conference on Pattern Recognition*, Vol.2, pages 137-140.

Lüdtke, N., Wilson, R.C. and Hancock, E.R. (2000a). Population codes for orientation estimation. In *Proceedings of the 15th International Conference on Pattern Recognition.*

Lüdtke, N., Wilson, R.C. and Hancock, E.R. (2000b). Tangent fields from population coding. In *Proceedings of the IEEE International Workshop on Biologically Motivated Computer Vision (BMCV2000)*, volume 1811 of Lecture Notes in Computer Science, pages 584-593.

Lüdtke, N. and Hancock, E. (2001). Gabor population codes for orientation selection. In Musio, C., editor, *Vision, The Approach of Biophysics and Neurosciences. Proceedings of the International School of Biophysics*, volume 11 of Series on Biophysics and Biocybernetics, pages 437-440, Casamicciola, Naples, Italy. Istituto Italiana Per Gli Studi Filosofici, World Scientic.

Lüdtke, N., Wilson, R.C. and Hancock, E.R. (2002a). Population coding of multiple edge orientation. In Dorronsoro, J., editor, *Proceedings of the International Conference on Artificial Neural Networks*, volume 2415 of Lecture Notes in Computer Science, pages 1261-1267.

Lüdtke, N., Wilson, R.C. and Hancock, E.R. (2002b). Probabilistic population coding of multiple edge orientation. In *Proceedings of the International Conference on Image*

Processing, volume 2, pages 865-868. IEEE.

Lüdtke, N., Wilson, R.C. and Hancock, E.R (in press) A Mixture Model for Population Codes of Gabor Filters. To appear in *IEEE Transactions on Neural Networks*

Acknowledgements

I would like to thank Prof. Edwin Hancock and Dr. Richard Wilson for their continuous support, strong encouragement, judicious criticism and accessibility. I am especially grateful for having been granted the freedom to incorporate biologically motivated techniques into my research.

I would also like to thank Dr. Adrian Bors for fruitful discussions, especially regarding the EM-algorithm, as well as my other colleagues from the computer vision group at York, for helping to create a friendly work atmosphere.

I wish to express my gratitude to Prof. Steven Zucker for hosting my research visit to Yale University–during which he provided invaluable insights into the nature of vision–and to the other members of the Yale computer vision group–especially Dr. Jonas August–for many crucial comments which proved to be instrumental.

I also wish to give a very special thank-you to the honourable Mrs. Antonia Johnson, who, through the William Gibbs/Plessey Award, provided the means for my visit to Yale.

Finally I would like to thank my parents for their generous support, and Mark Mahler for an unbiased perspective and his advice in revising the manuscript.

List of Figures

List of Tables

List of Symbols

θ	tangent/stimulus orientation
ψ	preferred orientation of a filter
r	(odd-symmetrical) filter response
$\bar{r}(\psi)$	expected response profile of a filter bank
G	Gabor response modulus
n	number of filters in a filter bank
$\vec{p}$	population vector
$p(\theta)$	pdf of local edge orientation
$p(r;\, \psi)$	pdf of a filter response value, given the preferred orientation ψ
m	number of mixture components in a von Mises mixture density
$P(i)$	i-th mixing coefficient
κ_i	concentration parameter of i-th mixture component
$\bar{\theta}_i$	mean of i-th mixture component
D_ν, ξ_ν	abbreviations for terms in EM update equations
$f(\theta)$	orientation tuning function of a filter
κ_0	concentration parameter of orientation tuning function
V_0	circular variance
ρ	resultant length of an angular pdf
γ	certainty measure based on Kullback-Leibler divergence
ϕ	angle of the line connecting two spline end points

$\mathcal{C}$	consistency between spline and actual intensity structure in an image
$\mathcal{G}$	complex Gabor filter
$\mathcal{I}$	Fisher information
$\mathcal{L}$	likelihood function
$I_0(x), I_1(x)$	modified Bessel functions of first kind (zeroth/first order)

Chapter 1

Introduction

The first step after data acquisition in many computer vision tasks is the extraction of local image features, such as edge orientation, texture, colour or shading flow. Their detection is notoriously ambiguous and error prone due to sensor noise and digitisation. Furthermore, they usually are strongly affected by clutter or change in illumination. Any efficient general purpose visual system should therefore represent the inherent uncertainty and ambiguity of sensory information at an early level, in order to avoid premature commitment to unreliable information. Local uncertainties should be resolved at later stages by relating local information to its context. This is clearly a non-trivial task since the number of possible combinations of features can easily lead to a "combinatorial explosion", where the search space becomes intractable. Thus two fundamental questions arise:

- How can local information be organised into robust global structures?
- How should local information be represented in order to support global organisation?

The objective of this thesis is to approach these complex problems from a biologically inspired point of view and to provide suggestions for overcoming them. Essentially

the novel contributions are the establishment of a theoretical link between the notion of population coding (a ubiquitous principle of distributed information representation in the brain) and the technical problem of object boundary detection, as well as a resultant algorithm that combines the detection of edges, corners and junctions into a unified probabilistic framework.

1.1 Perceptual Organisation and Gestalt Psychology

The idea that perception is a holistic process in which parts are linked according to their contextual relations was formulated during the 1920s by Wertheimer, Koffka, Köhler and other psychologists (Ellis, 1967) as part of "Gestalt" theory, whereby they established a number of "laws" of perception (see Figure 1.1) based on psychophysical investigations since the late 19th century.

The "Gestalt laws" are principles of perceptual organisation, a phenomenological description of the inbuilt bias in human vision by which local image elements are grouped and ambiguities are resolved. For instance, in Figure 1.1 (d) the scene is interpreted as two *overlapping* shapes with smooth boundaries. The alternative interpretation, two *adjacent* objects with more complicated outlines, is rejected despite both interpretations being *a priori* equally probable. In our natural environment, we are usually unaware of the active nature of our visual sense. However, specially designed artificial images, such as those in Figures 1.2 and 1.1, reveal that contour perception is more than a merely passive recording of sensory input (Kanizsa, 1979). It involves the binding of local features into perceptually salient groups ("pop-out"), completion of disrupted contours ("filling-in") and, as a result, segmentation. The optical illusions (Fig. 1.2) also illustrate that the perception of object outlines is not based on local

edge contrast alone, or else one would not see illusory or "amodal" contours (Kanizsa, 1979) where no actual intensity gradient (edge contrast) is present.

While it is often assumed that completion of large-scale gaps caused by occluding objects and "filling-in" of local contour disruptions resulting from low signal-to-noise-ratio can be accomplished within the *same* computational framework, it is important to distinguish between the two, since they actually require *fundamentally different* processing strategies (August and Zucker, 2000). Closing the large gaps in contours fragmented by an occluding object may even involve representing the occluder, requiring a more global degree of perceptual grouping.

This thesis will focus on short-range "filling-in" and contour grouping based on the Gestalt principles of "good continuation" and the closely related "good form" (Fig. 1.1 (c) and (d)). Though Gestalt psychology provides a merely *descriptive* theory with no direct clues as to how the computer scientist might incorporate perceptual organisation into an artificial vision system, principles such as "good continuation" can be expressed in geometrical terms (Kellman and Shipley, 1991), as will be explained in the literature review. Additional insight is provided by recent biological studies that have begun to shed light on the underlying physiological mechanisms of perceptual organisation at the single cell level.

1.2 Biological Fundamentals of Perceptual Organisation in Early Vision

In visual neurobiology, the concept of local feature detectors as elements of contour perception applies only to a simple isolated stimulus, such as an oriented bar. The response behaviour of so-called "simple cells" in striate cortex (V1), discovered by Hubel and Wiesel (1962), has been found to be less strictly linear than was origi-

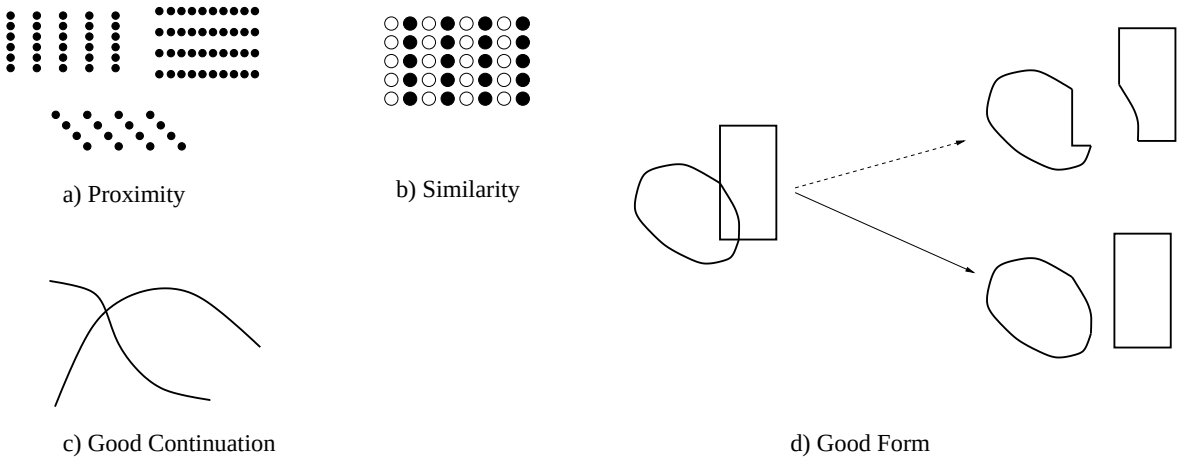


Figure 1.1: Examples of Gestalt principles. (a) Dot patterns are grouped with respect to proximity. (b) If the distance between dots is constant, elements are grouped with respect to other cues–in this case contrast. (c) The linear pattern tends to be interpreted as two smooth, intersecting curves rather than adjacent curves with tangential discontinuities. (d) The upper right interpretation–two adjacent forms with complicated outlines–is rejected in favour of the lower–two overlapping forms with simple and consistent boundaries–though both interpretations are *a priori* equally probable. Examples (c) and (d) are closely related, and both reveal a perceptual bias in favour of overlapping, rather than adjacent, structures.

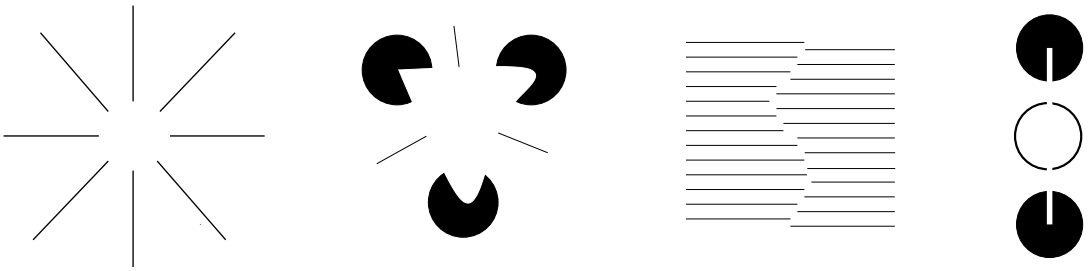


Figure 1.2: Examples of illusory contours. (a) Ehrenstein-illusion. In the centre a circle can be seen which appears brighter than the white of the background. (b) Modified Kanizsa-triangle. Instead of three "pacmen" and three line segments, a triangular shaped object seems to occlude three black disks and a linear structure in the middle. Again the object in the "foreground" appears brighter. (c) Gratings of abutting lines create the illusion of a curved smooth edge. (d) A bright white bar seems to overlap two disks and a circle. As in (a) and (b), there is no actual difference in brightness in the image.

nally assumed. For instance, neural firing rates saturate as stimulus contrast increases, a process which can be understood as a normalisation of neural responses within a hypercolumn (Heeger et al., 1996). Also, when stimuli of greater complexity are presented that exceed the size of the receptive field of an individual cell, *neighbouring* cells respond as a consequence of their retinotopic arrangement.

Due to various lateral connections between cells, different parts of a complex stimulus act as *contextual surround* for one another, and the principle of linear superposition is invalidated. The surrounding context can either facilitate or suppress neural responses, depending on relative contrast and orientation between centre and surround (Polat et al., 1998), and can even change the preferred orientation (Gilbert and Wiesel, 1990), perhaps explaining the Zöllner/"tilt" illusion (Fig. 2.2 in the next chapter).

Using stimuli such as those in Figure 1.2, neurons at early stages (V1 and V2) of visual processing have been found that respond to illusory contours (von der Heydt

and Peterhans, 1989a; von der Heydt and Peterhans, 1989b; Grosof et al., 1993). The conclusion of these experiments is that the perception of real and illusory contours is performed in the same area of the visual cortex at a low functional level.

In summary, cortical contour processing is very intricate and, despite the vast amount of known details, a concise physiological theory of perceptual grouping remains elusive.

1.3 Perceptual Organisation in Computer Vision

Perceptual organization in human and animal vision is a consequence of an adaptation to the *natural* environment. Erroneous perception of artificial images, such as the Zöllner illusion (Figure 2.2), is the price for enhanced performance in the processing of natural images (a mere subset of all possible images), which have particular statistical properties (Field, 1987; Krüger, 1998) that can be exploited in order to overcome the uncertainties of local image features.

Such features, which form the basis of natural and artificial vision, often have a small signal-to-noise-ratio, especially in "cluttered scenes", where object-background contrast is low. Nonetheless animals and humans cope amazingly well with such local distortions and their recognition capabilities are barely affected. It is difficult to imagine how an artificial system could yield comparable performance without incorporating some kind of technical equivalent to the Gestalt laws.

To overcome local distortions in machine vision, numerous authors (e.g., Lowe, 1985; Sha'ashua and Ullman, 1988; Heitger and von der Heydt, 1993; Sarkar and Boyer, 1994; Elder and Zucker, 1996; Guy and Medioni, 1996) have therefore proposed incorporating contextual relations among local features by combining responses of neighbouring feature detectors into a globalised and consequently more robust processing.

1.4 Biological Plausibility of Computer Vision Algorithms

"Biological plausibility" will here refer to the drawing of functional analogies between information processing in biological and technical systems, *not* to biological modelling.

In attempting to implement a biologically plausible machine vision system, the researcher combines methodologies from science *and* engineering, since the goal is to identify fundamental principles of perception relevant for both natural and artificial vision, and to separate them from computationally irrelevant physiological details.

The functional architecture of the visual cortex is, basically, determined by two factors: the statistics of the sensory input and the constraints imposed by the physico-chemical properties of cortical tissue. For instance, neurons in the central nervous system are mortal without the possibility of re-growth. They are also very noisy and, as a result of membrane capacities, have large time constants compared to electronic components. (The time scale of neural processing is within the range of milliseconds, whereas in electronics it is within nanoseconds.) Consequently, neural information processing must be highly parallel and robust against the failure of single neurons. In fact, a gradual decrease in performance, due to loss of neurons or signal distortions, has been identified as a key feature of distributed neural processing and is often referred to as the *principle of graceful degradation* (Rumelhart et al., 1986).

In computer vision, only those aspects of cortical information processing that are a consequence of adaptation to the structure of visual input (and the structure of the world), not physico-chemical constraints of biological neurons, need to influence technical implementation. It is argued in this thesis that the distributed nature of neural processing is not simply a necessary compensation for the shortcomings of individual neurons, but also an efficient coding strategy for local feature extraction in general.

Throughout this thesis, feature extraction will be performed by means of a bank

of Gabor filters, clearly a considerable abstraction from the ensembles of "simple" and "complex" cells in mammalian visual hypercolumns. No attempt will be made to capture the intricacies of intra-columnar connectivity in real cortical hypercolumns. Instead, a purely linear model of local feature extraction will be used that accurately describes the response behaviour of a linear Gabor filter bank.

1.5 Distributed Coding: a Paradigm for Local Feature Extraction in Machine Vision?

As previously mentioned, perceptual organisation of local information requires some flexibility in the representation of locally extracted features in order to avoid premature commitments at an early stage. A principal intention of this thesis is to demonstrate that distributed coding, a concept now generally accepted in neuroscience (Georgopoulos et al., 1986; Vogels, 1990; Young and Yamane, 1992; Wilson and McNaughton, 1993; Lehky and Sejnowski, 1990), is perfectly suited for this purpose, and can also be applied in *artificial* visual systems (in this case, to a bank of oriented filters). Neither the neural ensemble in a hypercolumn, nor the filter bank, represent the local orientation of a stimulus through the activity of a single element, but rather through an *activity profile* of the population as a whole.

Even the unresponsive units characterise the stimulus, in the sense that the concentration of activity in the population (loosely speaking the ratio of responsive to unresponsive units) describes the certainty of the encoded information. Expressed in more precise mathematical terms, a quantitative description of certainty yields a valuable additional piece of information for the selection of fiducial features, as will be demonstrated. It will also be shown that, in a perceptual grouping framework, the certainty of a local edge feature can be used to control the degree to which it is allowed

to be modified, during an updating step, in order to increase mutual consistency with neighbouring features.

1.6 Outline of the Thesis

Chapter 2 provides an overview of relevant literature on population coding and perceptual grouping focusing on edge/line continuation.

In Chapter 3, the population vector concept is applied to the problem of edge and line detection with Gabor filters. Also, the notion of certainty in population coding is explained and its role in the decision about the presence or absence of features is discussed.

The population vector algorithm is extended, in Chapter 4, to a probabilistic population coding approach. Edge detection is performed within a statistical framework based on an analytical model of the filter responses. By means of a version of the EM-algorithm (Dempster et al., 1977), a parametric model probability density function of local edge orientation is decoded from the Gabor filter bank.

In Chapter 5, the local population codes of tangent orientation, derived in Chapter 4, are combined in order to determine the parameters of splines interpolating between pairs of feature points, thus establishing a form of perceptual organisation in the sense of the Gestalt principle of "good continuation".

Chapter 6 summarises the results of the thesis and provides suggestions for future work.

Chapter 2

Literature Review

The purpose of this chapter is to review the relevant literature on population coding and perceptual grouping of contour segments, as separate fields of research. A *synthesis* of the population coding paradigm with the problem of contour detection is achieved through a novel method of contour detection, which will be introduced in the main body of the dissertation.

Part I: Population Coding

Through biological studies of various brain regions, it became evident during the 1980s that sensory, as well as motor, variables are represented by neural ensembles, rather than individual neurons, as had been postulated by Barlow (1972). The distributed coding strategy is often referred to as "population coding".

Typically, a neural population encoding one or several parameters consists of neurons exhibiting a rather broad tuning (Fig. 2.1). Consequently, a single neuron provides merely a course estimate of the encoded variable(s). However, by combining the outputs of ensemble members, the population as a whole is capable of accomplishing a substantially higher degree of representational accuracy. Due to its intrinsic redundancy, such a distributed code is also robust against neural noise and failure of individual units.

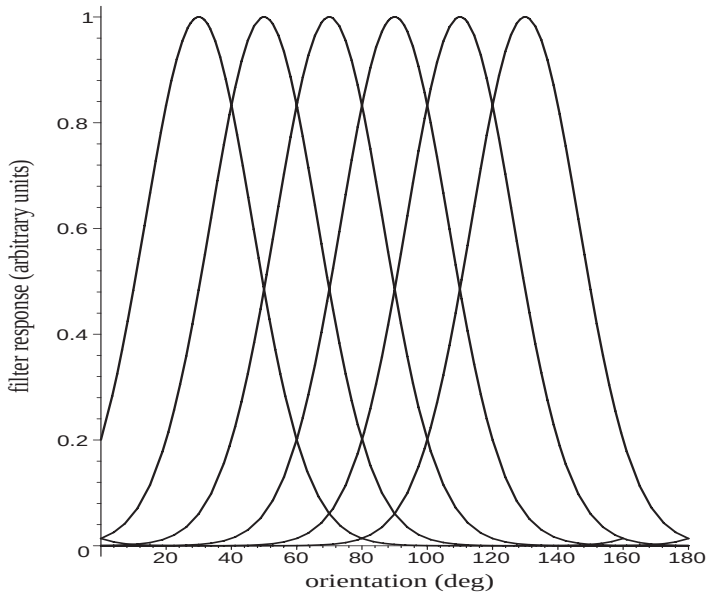


Figure 2.1: A set of bell-shaped, overlapping tuning curves covering a range of orientations create a redundant code that is robust against failure of individual units and noise. How can a precise orientation estimate be obtained from a population of such broadly tuned units?

Georgopoulos and colleagues (1986) have demonstrated that the direction of arm movements can be decoded from neural firing rates in the motor cortex of monkeys. Wilson and McNaughton (1993) have shown that the position of rats in their environment is encoded in populations of hippocampal neurons. Recording from these so-called "place cells", the authors were able to predict the position of the animals. Moreover, evidence has been provided by Young and Yamane (1992) of a face-encoding population in a region within the infero-temporal cortex, a brain area assumed to be involved in visual processing at an intermediate level between simple features and object representations. In the vision domain, population coding has also been linked to the perception of stereo disparity (Lehky and Sejnowski, 1990) and orientation estimation (Paradiso, 1988; Gilbert and Wiesel, 1990; Vogels, 1990).

Population coding has thus emerged as an essential paradigm in computational neuroscience. It is increasingly studied among theorists in the neural network community,

where research is mainly concerned with modelling biological information processing, rather than technical applications such as machine vision.

One principal issue in the computational neuroscience literature is coding accuracy and its dependence on the intrinsic noise in a neural network, as well as on the shape of neural tuning curves (e.g., Baldi and Heiligenberg, 1988; Zhang and Sejnowski, 1999), the number of neurons, and their distribution in parameter space (Vogels, 1990).

Paradiso (1988) was one of the first theorists to analyse the performance of a model hypercolumn of broadly tuned neurons by means of statistical estimation theory. His model offers an explanation for the high accuracy of discrimination in various sensory modalities, known as hyperacuity. For instance, the "just noticeable differences" observed in psychophysical orientation discrimination tasks are up to two orders of magnitude smaller than the orientation tuning widths of individual neurons (Vogels, 1990).

Population coding has also recently been extended to extracting probability densities from ensemble activities (Zemel et al., 1998), opening up the possibility of a more "holistic" processing of the information encoded in a population. In contrast to "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).

2.1 Population Vector Coding

There are several methods of decoding a population code. One is the so-called *population vector* or "centre-of-gravity decoding", a concept introduced by Georgopoulos and associates to describe, as mentioned, the representation of limb movements by direction sensitive neurons in the motor cortex (Georgopoulos et al., 1986). The basic idea is as follows.

Given a population of n neurons, each neuron is assigned a component vector with a magnitude proportional to the strength of its response, r_i (spike frequency), and a direction according to its preferred direction, denoted by the component vector $\vec{b}_i$. The vector components are then added, and the resultant vector sum is the population vector. Thus the population vector is a response-weighted average of n component vectors, $\vec{b}_i$, pointing in the neurons' preferred directions:

$$\vec{p} = \sum_{i=1}^{n} r_i \vec{b}_i \,. \tag{2.1}$$

If a vector $\vec{x}$ is properly encoded by the population, then

$$\vec{p} = \vec{x} \,.$$

It has been shown (Sanger, 1994; Sanger, 1996) that this equality implies cosine-tuning; i.e., responses are given by the dot product

$$r_i = \vec{x} \cdot \vec{b}_i \,, \tag{2.2}$$

which can be seen by inserting equation (2.2) in (2.1) and assuming the equality $\vec{p} = \vec{x}$:

$$\vec{p} = \sum_{i=1}^{n} (\vec{x} \cdot \vec{b}_i)\, \vec{b}_i = \vec{x} \left(\sum_{i=1}^{n} \vec{b}_i \, \vec{b}_i^{\,T} \right) \stackrel{!}{=} \vec{x} \,.$$

This yields the following condition for the outer product of component vectors:

$$\sum_{i=1}^{n} \vec{b}_i \, \vec{b}_i^{\,T} = \mathbf{1} \,, \tag{2.3}$$

where $\mathbf{1}$ denotes the unit matrix. The requirement (2.3) is always fulfilled with orthogonal basis vectors, i.e., when the component vectors are simply the Euclidean unit vectors, $\vec{e}_1, \ldots, \vec{e}_n$, but it imposes a severe restriction in the general case of an overcomplete basis $\{\vec{b}_i \ldots \vec{b}_n\}$. Any uneven distribution of preferred directions can easily violate

condition (2.3) and introduce a stimulus dependent bias to the population vector estimate (Salinas and Abbott, 1994; Sanger, 1996; Snippe, 1996). Though important for biological systems, such a bias is not a serious concern in a technical system, where one has complete control over the positioning of tuning curves in the parameter space.

2.1.1 Vector coding in biological and artificial vision

Vogels (1990) examined population vector coding of visual stimulus orientation in a model of striate cortical cells. The model is able to explain just noticeable differences of orientation discrimination in primate vision. Employing a sufficiently high number of units (several hundred), the obtained coding accuracy is less than 1°, even using tuning widths of up to 80°.

Gilbert and Wiesel (1990) used a very similar approach to explain the context dependence of orientation measurements and related it to physiological data and to the psychophysical phenomenon of "tilt illusion" (see Fig. 2.2). They investigated how a population vector estimate is affected by modifications of the ensemble activity profile induced by surrounding stimuli via long-range interactions between neighbouring populations. Such complicated *lateral interactions* not only alter the weighting of individual neurons but can even change their orientation tuning.

Expressed in the notation of the previous section, lateral interactions cause an alteration of the responses, r_i, which act as weights in equation (2.1), as well as a modification of the basis vectors $\vec{b}_i$. Since the task is merely to estimate the *direction* of the stimulus $\vec{x}$ (requirement $\vec{p} \| \vec{x}$ instead of $\vec{p} = \vec{x}$), condition (2.3) changes to:

$$\sum_{i=1}^{n} \vec{b}_i \, \vec{b}_i^{\,T} \propto \mathbf{1} \, .$$

Even this more relaxed constraint on the basis vectors can be violated by long-range interactions, which cause erroneous orientation estimates in certain geometric arrange-

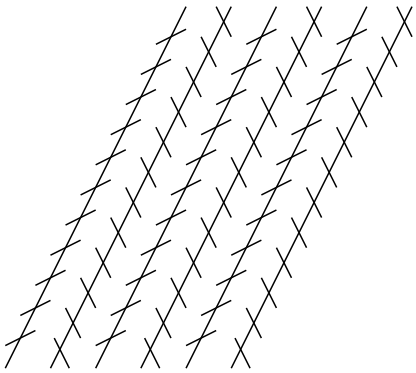


Figure 2.2: The Zöllner illusion ("tilt" illusion). The long lines are parallel but appear tilted. The small line segments influence the perception of orientation, which, according to Gilbert and Wiesel (1990), can be explained by the altered activities in the hypercolumnar ensembles caused by long range interactions between neighbouring orientation sensitive cells.

ments of linear stimuli, such as in Figure 2.2.

In computer vision, Granlund and Knutsson (1995) have defined a vector representation of local edge orientation, which the authors call "double angle representation". Based on a fixed set of four fundamental orientations (horizontal, vertical and the two diagonal orientations), a vector sum is computed. However, no connection with population vector coding has been established.

2.2 Maximum Likelihood Decoding

An alternative to vector decoding is maximum likelihood estimation (MLE). Based on Gaussian tuning functions and neural activities r_i governed by a Poisson distribution, a conditional density, $p(x|r_i)$, in the encoded variable is obtained using Bayes' law (Sanger, 1996). Assuming independent neural firing rates, the total likelihood of the

encoded parameter value is given by the product of the individual conditional densities:

$$\mathcal{L}(x|\mathbf{r}) = \prod_i p(x|r_i)\,. \tag{2.4}$$

Sanger (1996) has shown that equation (2.4) is essentially a product of tuning curves. Consequently, the likelihood function can be very sharply tuned, even for broad tuning curves.

MLE is effectively a form of template matching. The response data are matched with a template derived from the average population activity (Pouget et al., 2000). Since the method makes explicit use of the tuning function, it is not restricted to cosine-tuning implicitly assumed in the population vector approach (Oram et al., 1998). Theoretically, maximum likelihood is an optimal estimator; i.e., it is unbiased ("correct on average") and has the lowest possible variance, as determined by the Cramer-Rao bound (Deco and Obdradovic, 1996).

Though MLE may seem biologically implausible, Pouget and colleagues have shown that it can be closely approximated by recurrent biological networks (Deneve et al., 1999) for a wide range of tuning widths and input contrast values.

2.3 Extracting Probability Densities

The methods described previously aim to extract a *unique value* of the encoded quantity. However, the aforementioned MLE method proposed by Sanger (1996) 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, $\mathbf{r}$. (The response vector $\mathbf{r}$, i.e., the vector of ensemble activities, must not be confused with the population vector $\vec{p}$, which is the result of vectorial averaging of activities; see Section 2.1). Recovering a probability density is desirable because it maintains a distributed information representation, thus avoiding a premature commitment to local features that often have a high degree of

uncertainty. Additionally, a probability density can quantitatively characterise such inherent uncertainty of the estimate through its variance or a function thereof.

Zemel and colleagues (1998) have shown, however, that Sanger's "standard Poisson model" is incapable of representing densities broader than the tuning curve, a consequence of the multiplication of tuning curves described in Section 2.2. Moreover, since in Sanger's model $p(x|\mathbf{r})$ is always unimodal, it is impossible to represent ambiguity in x.

Both limitations in representational capacity are overcome in a recently developed, more refined probabilistic model by Zemel and co-workers that is able to decode multimodal distributions of arbitrary width. Their distributional population coding approach, the "extended Poisson model", is a non-parametric method which aims to recover the encoded probability density as a set of sampling points. The discrete values of the probability density are themselves treated as stochastic variables and are determined by maximising the data likelihood function. Thus the algorithm performs an approximated form of maximum a posteriori estimation in *distributions over distributions*.

As in Sanger's model, a neural response r_i to a stimulus x is governed by a Poisson distribution p(r) around a mean determined by the corresponding tuning curve $f_i(x)$. If the encoded variable is governed by a probability density $p(x)$, the average neural activities are given by:

$$\bar{r}_i = \int_X f_i(x)\, p(x)\, dx\,, \tag{2.5}$$

which is the fundamental encoding equation. However, in order to incorporate arbitrary probability distribtions of the stimulus, p(x), both the p(x) and the tuning function are approximated by piecewise constant histograms. Let $\hat{\phi}_j$ be the value of the pdf $p(x)$ in the interval $(x_j, x_{j+1}]$ and f_{ij} the value of the i-th tuning function in $(x_j, x_{j+1}]$. The average ensemble activities are then expressed by a discrete approximation of the

encoding equation (2.5):

$$\bar{r}_i = \int_X f_i(x)\, p(x)\, dx \approx \sum_{j=1}^{M} f_{ij} \hat{\phi}_j \,.$$

The unknown sampling points $\hat{\phi}_j$ of the wanted pdf p(x) are themselves treated as random variables. The task is then to perform maximum a posteriori (MAP) estimation in order to find the $\hat{\phi}_j$, i.e., to maximise

$$\log p\Big[\{\hat{\phi}_j\} \Big| \{r_i\}\Big] \quad \text{under the constraint} \quad \sum_{j=1}^{M} \hat{\phi}_j = 1 \,.$$

The number of sampling points ("histogram bins"), M, exceeds the size of the population, and regularisation is accomplished by means of a smoothness prior that restricts the change $\hat{\phi}_{j+1} - \hat{\phi}_j$ of the pdf in neighbouring points.

One possible means to estimate the $\hat{\phi}_j$ is simple gradient descent. In fact, Zemel and colleagues reformulate the MAP estimation as a maximum likelihood estimation by interpreting $\sum_{j=1}^{M} f_{ij}\hat{\phi}_j$ as a mixture model where the $\hat{\phi}_j$ play the role of mixing coefficients. A version of the EM-algorithm is then used to determine the MAP values of the $\hat{\phi}_j$.

2.4 Complex Feature Detectors

An issue mentioned only briefly in the population vector model of Gilbert and Wiesel (1990) is complex stimuli *within* the classical receptive field. Two-dimensional structures with multiple orientation create multimodal response activity profiles in the hypercolumnar ensemble. How the brain processes such stimuli is unclear, since experimental data is scarce.

This matter is also highly relevant in machine vision, where an analogous situation arises when, in addition to regular edges, corner points and junctions are to be detected. In the presence of two-dimensional intensity features, classical gradient

based edge detection schemes are inefficient 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 specially designed corner detectors. However, the biological plausibility of such detectors is strongly debated.

While some researchers (Heitger et al., 1992) regard the so-called end-stopped cells as detectors of occlusion points (T-junctions), their assumption has been questioned due to the varying and often very limited degree of end-stopping that such cells actually exhibit. For example, Zucker and co-workers interpret end-stopped cells as curvature detectors (Dobbins et al., 1987).

Zetzsche and Barth (1990) have identified fundamental limitations of linear filters operating on two-dimensional intensity features. Typically, a linear filter designed to detect 2D-features involves a combination of responses from linear components. For signal-theoretic reasons, Zetzsche and Barth argue that 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 which the linear components are combined.

In accordance with Zucker's viewpoint, it is argued in this thesis that a hypercolumnar ensemble, respectively a filter bank, is capable of representing complex local intensity structure, excepting curvature, and that no additional corner detector is required.

Using their distributional population coding model, Zemel, Dayan, and co-workers have shown how multiple motions (Zemel and Dayan, 1999) and multiple orientations (Zemel and Pillow, 2000) can be decoded from biological neural networks. Referring to Zemel and colleagues, the problem of representing points of multiple edge orientation will be addressed in Chapter 4 as a principal short coming of the population vector

approach, and a solution will be offered within a framework of probabilistic population coding adapted to machine vision. The issue of response ambiguities, raised by Zetzsche and Barth (1990), will also be resolved in this context, since probabilistic population decoding involves non-linear operations in the form of products of functions of ensemble activities.

2.5 Combining and Transforming Population Codes

Another topic in the literature is the transformation and combination of population codes in order to encode complex quantities that transcend the sensory and/or motor variables encoded by the original populations.

One example of a transformed population code is colour vision. As pointed out by Lehky and Sejnowski (1998), the retinal colour population code consists of neurons tuned to optical wavelength, but the combined neural activities can represent the percept "white", which itself is *not* a wavelength. Combined neural populations can span abstract representational spaces comprising dimensions that, though derived from sensory and motor inputs, have no direct physical counterpart in the outside world (Zemel and Dayan, 1997; Deneve et al., 2001).

In vision, the information of neural populations with highly localised receptive fields, such as the simple and complex cells in V1, is combined to form neural populations at an intermediate level in the visual pathway that are responsive to complex stimuli of large spatial extent, as observed in the inferotemporal cortex (Tanaka, 1996).

2.6 Conclusion

The main population coding concepts from computational neuroscience have been reviewed and their advantages and drawbacks have been discussed. The ubiquity of population coding in highly efficient biological sensory systems provides the motivation to investigate the application of population coding in image processing pursued in this dissertation. The relation to concrete applications in edge detection will be substantiated in chapters 3 and 4. In chapter 5 it will be shown how a combination of local population codes of edge orientation can lead to perceptual organisation of contour segments.

The essential feature of population coding is its great flexibility and representational capacity. For the purpose of this thesis, the probabilistic approach provides the means to accurately represent local contour orientation and to characterise the reliability of orientation estimates through the inherent uncertainty encoded by the corresponding angular probability density.

Literature Review Part II: Perceptual Grouping

Since the literature on perceptual organization is extensive, it is beyond the scope of this thesis to cover all the existing methods. Therefore the purpose of the following sections is to review some important papers representative of different approaches to perceptual organization of contour features, and to describe the performance of these algorithms on real images.

2.7 Contour Integration And Perceptual Saliency

An important function of perceptual grouping is the restoration of contours, since edge information extracted by means of local operators is inevitably fragmented. Each feature gives a local value of confidence that an edge segment is present, and possibly additional information about its orientation and curvature. Sometimes, due to noise and clutter, the edge contrast is very poor or varies strongly along a contour. The task of a data driven grouping algorithm is then to relate edge fragments and to enable mutual support between individual local features based on their directional consistency, thus enabling the transition from local to global features, i.e, from edge segments to straight lines or smooth curves. The process is closely related to finding and enhancing regions that are *perceptually salient*; i.e., a technical analogue of the "pop-out" effect

in psychophysics.

2.7.1 The Saliency Network of Sha'ashua and Ullman

Sha'ashua and Ullman (1988) defined a measure of perceptual saliency of a curve, based on geometric properties. The saliency measure increases monotonically with the length of the evaluated curve and decreases with its total squared curvature. Additionally, the degree of fragmentation, expressed in terms of the number of gaps and total gap length, is penalised. Using the saliency function, a "saliency map" is defined as an image in which the intensity encodes the saliency value of the most salient curve emanating from the point under consideration. The network performs a relaxation procedure to maximise the saliency measure.

In their critical assessment, Alter and Basri (1998) show that the saliency network has some shortcomings, producing a number of results that differ from human perception:

- A curve with one large gap is considered more salient than one with a few small gaps of equal total length.
- The network has difficulties representing more than the most salient curve.
- The saliency measure is not invariant with respect to uniform scaling of the image.
- Erroneous binding can occur through the merging of contour segments belonging to different objects.

2.7.2 Kellman and Shipley's Theory of Visual Interpolation

Kellman and Shipley (1991) developed a theory of visual interpolation based on psychophysical evidence. By introducing a geometrical definition of edge segment "relatability", they provide a precise mathematical description of the Gestalt law of "good

continuation". Figure 2.3 shows a geometric construction of the concept of "relatability", which accurately describes the bias in human perception towards smooth contours and which can be directly implemented in a machine vision system. Two edge segments (tangent segments) centred around points P_1 and P_2 are "relatable" if, and only if, the angle δ at the point of intersection of the edge extensions is acute, corresponding to an interpolating curve with a moderate degree of bending and without inflection points.

In practice, however, the geometrical configurations of edge segments obtained from an edge detector are often not as precise as required, due to noise. As a result, two edge segments belonging to the same contour can appear "unrelatable", even if their orientations deviate only slightly from the true values. This is expected to happen frequently with straight contours, since image noise creates an angular "jitter" in the tangent orientations measured along the contour.

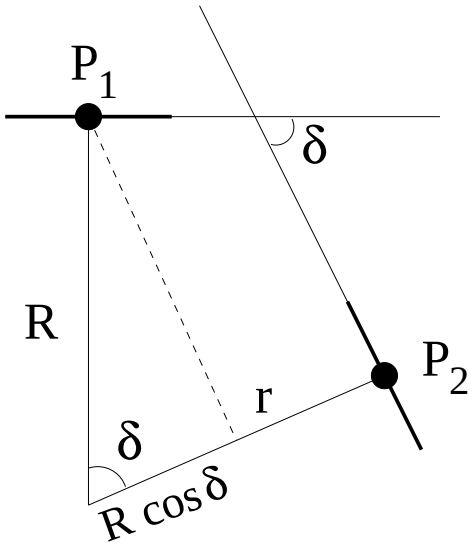


Figure 2.3: A construction defining the concept of "relatability" according to Kellman and Shipley (1991). Two edge segments are relatable, if and only if, $0 \leq R\cos\delta \leq r$. Consequently, $0 \leq \delta \leq \pi/2$, which excludes interpolation between edge segments that are parallel or require a connecting curve with sharp bending ($R\cos\delta < 0$).

2.7.3 Global Contour Grouping: The Extension Field

Guy and Medioni (1996) devised an algorithm for contour grouping based on the Gestalt principles of co-curvilinearity ("good continuation") and proximity. After applying an edge detector, a convolution is performed on the edge map using a special mask called *extension field*, a vector field encoding the likelihood and orientation of possible continuations from an edge segment at its centre to all other points in the image. Thereby a large number of smooth curves are represented simultaneously. The direction of the extension field at a point (x, y) equals the tangent angle of the most likely curve connecting (x, y) with the edge segment at the centre of the extension field. The magnitude of the vector field in (x, y) is the likelihood of the existence of the connecting curve.

Referring to Sha'ashua and Ullman (see Section 2.7.1), the extension field is defined under the assumptions that human contour grouping favours curves of minimum total curvature and that the completion of fragmented contours is accomplished in terms of constant curvature. Therefore the tangent orientation at a particular point is chosen to be that of the *osculating circle* connecting the point and the central edge segment. By superposition of extension fields emanating from all edge locations, every pixel receives a number of vector "votes". This process is a vector convolution resulting in a two-dimensional covariance matrix for each point. Then the principal axis of all the orientation votes is taken as the edge orientation. The eigenvalues are interpreted as the axes of an ellipse, the eccentricity of which describes the degree of "agreement" among orientation votes, i.e., a measure of certainty. Multiplication of eccentricity by the strongest eigenvalue, which depends on the edge contrast, yields a *saliency* measure.

Since the extension field combines votes from all edge locations in the image, it is an implicitly global grouping scheme that differs from many previously developed techniques that use local operators to infer global structures. The system can operate

on tokens with different degrees of orientational certainty, from line segments with a precise orientation to points (complete uncertainty of orientation). A number of visual illusions can be reproduced, such as the Ehrenstein circle (see Fig.1.2 a) and a Kanizsa-type figure (see Fig.1.2b), but the performance on natural images is not demonstrated.

2.8 Algorithms Modelling Biological Data

Based on the neurophysiological experiments of von der Heydt and Peterhans (1989a, 1989b), Heitger and associates (1992, 1998) developed a model of illusory contour perception in mammalian visual systems focusing on the close relationship between illusory contours (such as in Fig. 1.2) and occlusion.

The essential ingredients are models of neurons found in areae V1 and V2 of the visual cortex. Complex cells are modelled as the moduli of complex Gabor filters and referred to as C-operators. The so-called "end-stopped cells" (cells responding to sudden terminations of contours, interrupted contours or corners) are modelled as combinations of two or three C-operators.

A combination of end-stopped operators then serves as a tool to find the "key points" of occlusion, i.e., locations where contours terminate due to occlusion. Because of the special geometric arrangement of operators, corner points can be excluded.

Once the key points of occlusion are found, the map of end-stopped operator responses at key points is convolved with a "*grouping field*", an oriented club-shaped kernel. The result of the convolution is added to the C-operator responses, whereby contours of weak or vanishing contrast are restored.

The authors define two types of grouping, "ortho-grouping" and "para-grouping", meaning that the restored contour and the endpoint termination are either orthogonal or aligned, respectively. Due to the arrangement of end-stopped operators in the

ortho- and para-grouping scheme, it is possible to infer some information about the foreground/background structure of the overlapping contours in the image.

The model seems to perform well on natural images, apart from some ambiguities in foreground/background segregation and occasional erroneous filling-in. The authors conclude (Heitger and von der Heydt, 1993) that these problems can only be overcome at higher stages of processing. The robustness against noise is not discussed.

2.8.1 Li's V1-Model of Contour Integration

Zhaoping Li's neural model of contour integration (1998) tries to explain the detection of perceptually important structures based only on elements of the primary visual cortex (V1). The model incorporates oscillating neurons sensitive to local stimulus orientation and inhibitory interneurons. Interacting via lateral intra-cortical connections, neurons of similar orientation preference can excite and synchronise one another's activity, or inhibit one another through interneurons. The model also seems to reproduce the effect of pop-out of contours from a noisy background, which accords with recent biological findings on the relation between lateral connections in the visual cortex and contrast dependent facilitation, or suppression, of oriented stimuli by their surrounding context (Stemmler et al., 1995). Stemmler and associates have suggested that neural responses corresponding to an unrelated or noisy background can suppress one another due to their lack of mutual consistency. The responses produced by smooth object contours are much more consistent, enabling mutual enhancement, and thus segregating figure and background. However, Li's model considerably simplifies the complicated dependencies of contextual facilitation and suppression between stimuli on relative contrast and orientation.

Though the model presents a *pre-attentive* mechanism, it also addresses the possible role of top-down feedback as an additional, though not necessary, element for the

purpose of contour integration. The possible feedback is not explicitly modelled, but it is shown how higher levels might control contour processing via projections on the inhibitory interneurons (negative feedback). A more direct influence of feedback signals on the excitatory direction sensitive neurons themselves is not considered. The reason for such a restriction is simplicity of processing in terms of stability, and the ongoing debate about the actual target neurons of feedback projections.

The model can enhance contour elements of weak contrast, but it does not explain illusory contours (vanishing edge contrast), as it relies exclusively on V1 elements. It is argued that, as opposed to contour enhancement, at least V2 and possibly even higher levels, as well as the feedback from these areae, are necessary to model the perception of illusory contours. The algorithm is tested on artificial images (line drawings) and one natural image, where it emphasises contours and reduces background noise.

2.9 Computing Contour Closure

Elder and Zucker (1996) address the problem of computing closed bounding contours. Their approach is motivated by psychophysical findings that closure seems to be a very important grouping cue which can override those of regional texture (Fig. 2.4). Therefore, no restrictions are placed on the type or shape of the bounded structures, taking into account the often highly inhomogeneous appearance of objects in natural images.

A multi-scale edge detection algorithm yields information about edge position and tangent orientation from which a sparsely connected tangent graph is constructed. Each node is assigned with the tangent information and, according to a Bayesian model of tangent linking, each arc is labelled with the likelihood that the corresponding tangent pair forms a contiguous component of the same contour. Each node is connected to only a small number of neighbours (usually six), according to the most likely pairings.

The goal of closure grouping is then to find the *maximum likelihood cycles* for every tangent in the graph. Thus the grouping task is reduced to a shortest path problem, which can be solved by means of standard techniques within a reasonable amount of time.

The performance of the algorithm is tested on natural images. Since no restrictions on type and shape of the bounded structure have been made, the algorithm is able to find boundaries of perceptually meaningful, but very heterogeneous, parts of an image. The result is an improvement over many other approaches that either fail to identify closed boundaries, since they focus on smoothness only, or, in the case of texture based segmentation, decompose objects into pieces, due to their inner heterogeneity. As with other data-driven algorithms, occasional binding errors cannot be avoided.

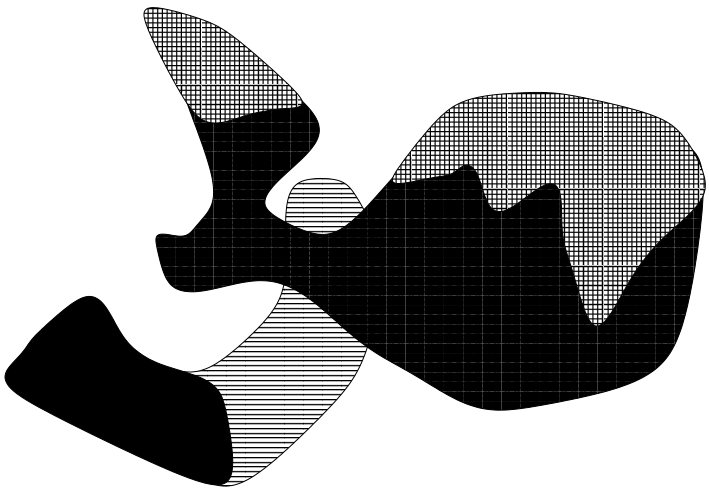


Figure 2.4: A synthetic image in which human observers perceive two overlapping objects. Region based grouping, however, would segment six disjoint parts. Since there are only "non-sense" objects, human segmentation cannot be based on familiarity. Elder and Zucker (1996) conclude that it is the geometry of the boundaries which guides our perception (illustration after Elder and Zucker).

2.10 Spline Representations

Because of their properties of smoothness, splines have been widely used for the representation of curvilinear object boundaries in computer aided design (CAD) for decades. Various types have been developed (Farin, 1988). While in CAD-applications the spline control points are usually determined by the user, a computer vision system should be able to find them automatically, based on an initial feature detection procedure, a problem addressed by Zucker *et al.* (1988), Cham and Cipolla (1996), Leite and Hancock (1997) and others.

Closely related forms of curves for detecting *closed* boundaries are the well-known "snakes", which are energy minimising splines that serve as a flexible contour template (Kass et al., 1988). The "snakes" realise a trade-off between smoothness constraint and matching of intensity features, such as edge segments. In practice, however, their initialisation proves difficult and often requires manual intervention. Another obstacle is the often highly complex topology of the intensity gradient in natural images, which makes it difficult to detect global contour features.

2.10.1 The Role of Curvature in Computer Vision.

In the context of splines as a means for contour representation, the question of the role of curvature arises. There is an old debate over the perceptual significance of curvature in perception. It has been argued (Attneave, 1954) that local extrema of curvature are the most perceptually important points on object contours, which in turn motivated Perez de la Blanca and associates to develop a spline-based algorithm for contour representation in a graph with landmarks at points of maximum local curvature (Perez de la Blanca et al., 1993). The points of highest curvature are found iteratively.

By contrast, Lowe (1985) argued that the perceptual significance of local extrema of curvature is commonly overestimated. Firstly, when three-dimensional curves are pro-

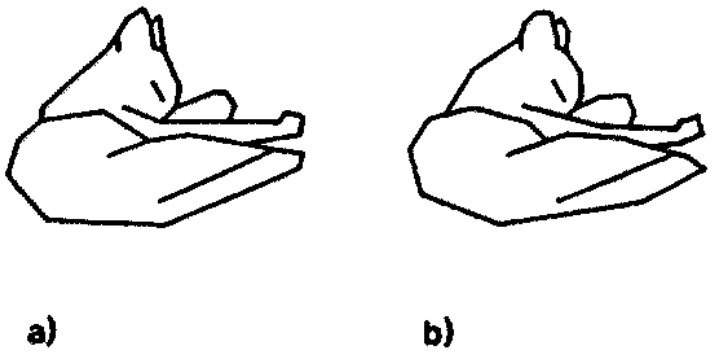


Figure 2.5: Two drawings derived from the same picture of a cat. (a) was created by connecting the points of local maximum curvature with straight lines; (b) by connecting points *between* those of maximum curvature (leaving line terminations at the same place). Both drawings are perceptually similar and easily recognisable, though the points in (b) are as far removed from those in (a) as possible, leading to the conclusion that the points of maximum curvature are not necessarily more perceptually significant (illustration from Lowe [1985]).

jected onto a two-dimensional image, the curvature maxima are not invariant, unlike curvature inflection points, discontinuities in tangent orientation and curve terminations. Secondly, it can be shown (Fig. 2.5) that in an "abstract" line drawing created from a picture by selecting certain points and connecting them with straight lines, the content is equally well represented, whether or not the points are curvature extrema or placed elsewhere on the contour.

Parent and Zucker pointed out that local ambiguities in contour detection can be resolved by means of curvature consistency, even though the measurement of curvature is more prone to error than that of local orientation (Parent and Zucker, 1989). The consistency of curvature seems to be more relevant than the exact value at specific locations (Zucker et al., 1989).

Krüger introduced a novel kind of wavelet-type filter as a generalisation of Gabor filters (Krüger and Peters, 1997). The sinusoidal part of the filter kernel and its Gaussian

envelope are both bent parabolically, rendering it sensitive to curved stimuli. Because of their curved shape, they have been given the humorous name "banana wavelets".

The features are organised in a metric space which contains the normal Gabor wavelets as the subspace of vanishing curvature. By means of the metric, a clustering of features is performed in order to extract the fiducial features from a training set of images showing examples of a certain class of objects (Krüger and Lüdtke, 1998). The result is a symbolic representation of the object class (Fig.2.6c). However, the curvature tuning of the wavelets is rather broad (Krüger personal communication), allowing only a coarse measurement of curvature.

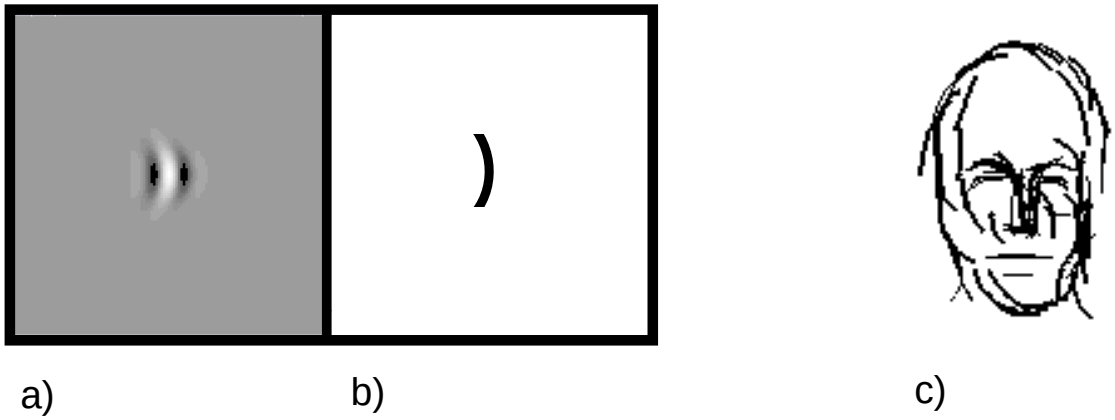


Figure 2.6: a) A Banana wavelet. b) The symbolic analogue, a curved line segment. c) A symbolic representation of the object class "face", from Krüger and Lüdtke (1998).

2.10.2 Token Based Grouping With Splines

Dolan and Riseman (1992) developed an algorithm for perceptual grouping of tokens such as straight line segments, corners, inflections, etc., represented by conic splines, which are curves in the form of conic sections, including hyperbolae and ellipses, in addition to the usual parabolae (Farin, 1988). The contour segments are assigned to nodes in a so-called link graph, and the arcs are labelled with the result of a binary compatibility relation among neighbouring tokens. The compatibility depends on distance,

difference of tangent orientation at the end points of tokens (angular compatibility), and the degree of overlap among tokens.

In an iterative procedure, tokens of sufficient compatibility can be replaced by a larger token, in such a way that the information about the constituting smaller tokens is not erased, resulting in a hierarchical contour representation across different scales that can be traced back to the lowest level. The parameters determining the compatibility function change with respect to scale, preserving high frequency structures (such as corners) on the fine scale, and reducing noisy fluctuations on larger scales by more rigorous smoothing.

2.11 Conclusion

After introducing the basic concepts of population coding, a number of perceptual grouping techniques have been reviewed. Though the two topics have thus far been dealt with separately, some relations between population coding and perceptual organisation have already become apparent. For instance, the population vector model of Gilbert and Wiesel (1990) investigates the influence of horizontal connections between local populations on the encoding of stimulus orientation. This interpretation of perceptual organisation as a combination of local population codes will provide the link between population coding and perceptual grouping in the thesis.

In many grouping algorithms the notion of certainty is neither made explicit, nor quantitatively represented. Confidence in the presence of a particular intensity feature is commonly identified with the response magnitude of a corresponding detector. In the case of edge detection, certainty of an edge orientation estimate is therefore closely linked with edge contrast (magnitude of the intensity gradient). Guy and Medioni's algorithm is the only reviewed method that treats certainty as a *contrast independent* quantity, though only at the grouping level.

Throughout this thesis, certainty will be considered at the level of local feature extraction. It will be demonstrated that, in the context of distributed coding, certainty can play a more distinct role when defined in terms of the spread of activity across the feature detector ensemble. For unimodal activity profiles, a measurement (in this case a local orientation estimate) is the first moment/mean of the ensemble activities, and the certainty of that measurement can be associated with the second moment/variance or a function thereof. A generalisation of this concept to multimodal response profiles will be provided within a probabilistic population coding scheme that decodes a mixture distribution from ensemble activities.

The combination of the overall strength of detector responses and the corresponding certainty can help level out the local variations of responses on contours and render local feature detection more robust against noise.

Chapter 3

Population Vector Coding

This chapter addresses the problem of local orientation selection (tangent field estimation). The intention is to demonstrate the practical applicability of population vector coding for this purpose and to relate computer vision experiments to the theoretical results in the computational neuroscience literature, as well as to compare the performance of the population vector to other methods of tangent field extraction and orientation estimation.

Following the work of Heitger, von der Heydt and associates (Heitger et al., 1992) and Lades et. al (1993), Gabor filters are employed as a simple mathematical model of orientation sensitive cells in the cortical hypercolumn. Adopting the biological concept of *population vector decoding* (Georgopoulos et al., 1986; Vogels, 1990), a continuous orientation estimate is extracted from the discrete set of responses in the Gabor filter bank by performing vectorial combination of the broadly orientation-tuned filter outputs. The result is a population vector, the orientation of which gives a precise and robust estimate of the local contour orientation. The accuracy and noise robustness of orientation measurement and contour detection is also investigated, as well as the relationship between the certainty of the orientation estimate and the shape of the response profile (ensemble activity) of the filter bank. Comparison with alternative

methods of orientation estimation reveals that the tangent fields resulting from the population vector technique provide a more perceptually meaningful representation of contour direction and shading flow.

3.1 Properties of Gabor filters

Gabor Filters have been proposed as a model of orientation selective cells ("simple cells") in the striate cortex (Marcelja, 1980; Daugman, 1985) and have found numerous applications in computer vision, including edge detection (Shustorovich, 1994), detection of line orientation and width (Buse et al., 1996), texture analysis and object recognition (Lades et al., 1993).

In the above-mentioned analogy to biological information processing, the real (cosine) part of a complex Gabor filter corresponds to simple cells of even symmetry and the imaginary (sine), to simple cells of odd symmetry.

3.1.1 Filter Kernels

A general complex Gabor filter of wavelength λ and preferred orientation θ is defined as

$$\mathcal{G}(\hat{x},\hat{y};\theta,\lambda) = \frac{1}{2\pi\sigma_w\sigma_e\lambda^2} \underbrace{\exp\left(-\frac{1}{2\lambda^2}\left[\frac{\hat{x}^2}{\sigma_w^2}+\frac{\hat{y}^2}{\sigma_e^2}\right]\right)}_{\text{Gaussian envelope}} \left\{ \underbrace{\exp\left(\frac{2\pi}{\lambda}\hat{x}\ i\right)}_{\text{sinusoid}} - \underbrace{\exp\left(-2\pi\sigma_w^2\right)}_{\text{mean-correction}} \right\}, \tag{3.1}$$

where $\hat{x}$ and $\hat{y}$ are the coordinates of an image point in the coordinate system rotated by the angle θ:

$$\hat{x} = x\cos\theta + y\sin\theta \qquad \text{and} \qquad \hat{y} = -x\sin\theta + y\cos\theta\,.$$

In (3.1), σ_w and σ_e are the width parameters of the Gaussian envelope in units of the wavelength, defined by the scale invariant expressions

$$\sigma_w = \frac{\sigma_x}{\lambda} \qquad \text{and} \qquad \sigma_e = \frac{\sigma_y}{\lambda}\,.$$

The index w stands for "wave", since σ_w determines the size of the envelope in the direction of the sinusoidal oscillation, and σ_e (as in "edge") denotes the width in the preferred orientation of the filter. Following Lades (Lades et al., 1993), the term $\exp(-2\pi\sigma_w^2)$ in equation (3.1) is subtracted, in order to obtain a filter of zero mean, ensuring independence of mean intensity. Heitger and associates (Heitger et al., 1992) have proposed frequency modulation to achieve independence of mean intensity, but the computation of the kernels then becomes more complicated.

Furthermore, it is useful to consider the *modulus* of the convolution of an input image I with the complex Gabor filter $\mathcal{G}$, i.e., the Pythagorean sum of even-symmetric and odd-symmetric filter responses, which will be denoted by G for simplicity:

$$G = |\mathcal{G} * I|\,.$$

The square of G is often referred to as *local energy* (Adelson and Bergen, 1985; Morrone and Owens, 1987). The properties of Gabor response moduli resemble those of "complex cells", since they do not distinguish contrast polarity (edges vs. lines) and are robust against small stimulus translations within the receptive field (Heitger et al., 1992; Lades et al., 1993; Heitger et al., 1998). The robustness of the response modulus has been successfully utilised in artificial face and object recognition, e.g., by Lades et al. (1993).

Throughout this chapter, only Gabor moduli will be considered. Additionally, all filters will have radial symmetry, i.e., $\sigma_e/\sigma_w = 1$.

3.1.2 Orientation Tuning of Gabor Filters

The essential element in the analysis of a population code is the tuning curve, in this case the *orientation* tuning curve, which describes the relation between filter response and the orientation of the local edge/line structure. The quantity that characterises the shape of the tuning curve is the tuning half width, w (i.e., the half-width at half height of the tuning curve), and it is important to know how w depends on the filter parameters.

Like simple and complex cells in striate cortex, Gabor filters have rather broad orientation tuning, which depends on an essential control parameter for the shape of the filter kernels, the width of the Gaussian envelope, denoted by σ_e. Though, in principle, the filter mask extends across the entire image, a value of three times σ_e can be considered the *effective radius of the "receptive field"* (radius of the filter mask), since the Gaussian envelope virtually vanishes at greater distances from the centre.

To determine the tuning curve and to examine the influence of the filter envelope width on the tuning width, synthetic images of straight lines were used as test stimuli. The line thickness in all experiments was one pixel, and orientations ranged from 0° to 170° at 10° intervals. Figure 3.1 shows the tuning curves for three filters with a preferred orientation $\psi = 90°$, a wavelength $\lambda = 8$ pixels, and envelope widths $\sigma_e = 0.6$ (4.8 pixels), 1.0 (8 pixels) and 2.0 (16 pixels), respectively. The estimated half-widths of the tuning curves are $w = 17.2°$, 9.7° and 5.2°. The first value is comparable to typical orientation tuning half-widths of striate cortical cells (Vogels, 1990). Interestingly, as will be demonstrated later, this proves to be the most suitable tuning width for orientation measurement.

The responses in Figure 3.1 are fitted with a model orientation tuning function. Here the von Mises function (Mardia, 1972) is used, since it is appropriate to angular variables and consistent with biological studies (Swindale, 1998). The normalised form

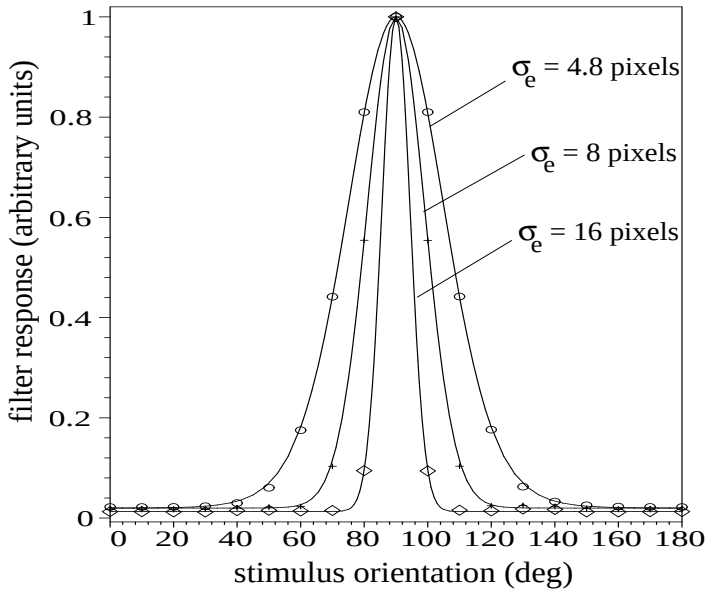


Figure 3.1: Normalised empirical tuning curve for the moduli of three Gabor filters ($\lambda = 8$ pixels, vertical orientation). The "stimuli" are single straight lines in 256×256 grey-scale images. Response magnitude is averaged over several points on the line, avoiding the region close to the end points. The half-widths are $w = 17.2°$, $9.7°$ and $5.2°$. Responses are fitted with von Mises functions (see equation 3.2).

of the tuning function is:

$$f(\theta;\psi) = \left(\tfrac{1-f_0}{e^{\kappa}-e^{-\kappa}}\right)\ e^{\kappa\cos[2(\theta-\psi)]}\ - \left(\tfrac{1-f_0}{e^{2\kappa}-1}\right) + f_0\,, \tag{3.2}$$

where ψ is the preferred orientation of the filter, κ is the so-called concentration parameter, playing a similar role as the inverse of the width σ of a Gaussian, and the offset f_0 is a "mean activity", modelling the effect of discretisation noise.

The function $e^{\kappa\cos[2(\theta-\psi)]}$ is always greater than zero. To ensure that f_0 is the *total* offset along the ordinate, the term $(1-f_0)/(e^{2\kappa}-1)$ is subtracted. Consequently, the tuning function has values in the range $[f_0, 1]$. Thus κ, ψ and f_0 serve as independent control parameters of the normalised model tuning function.

In Figure 3.2 the tuning width is plotted as a function of the kernel width σ_e. The two quantities are inversely proportional to one another due to the general uncertainty

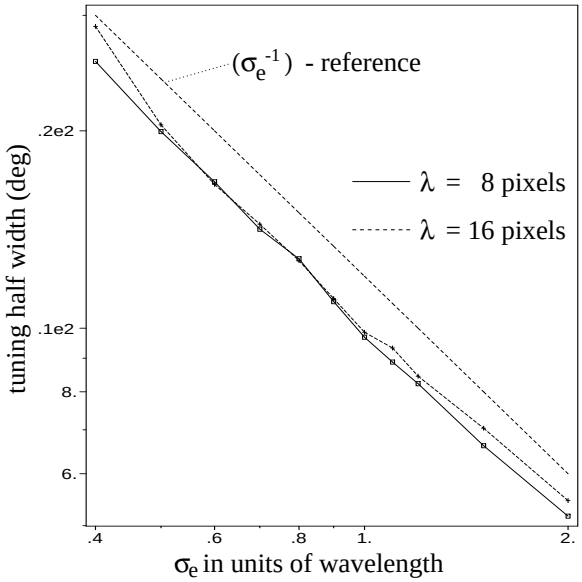


Figure 3.2: Log-log-plot of the tuning half-width as a function of the kernel width illustrates the uncertainty relation between tuning width and spatial width: $w \propto (\sigma_e)^{-1}$.

relation between orientational bandwidth (i.e., the tuning width w) and spatial width (σ_e). In fact, Gabor filters have been shown to minimise the quantity $w\,\sigma_e$ (Daugman, 1985).

The main conclusion to be drawn is that Gabor filter banks provide rather coarse estimates of feature orientation, unless the full range of orientations is sampled with a large number of filters, which would obviously be highly inefficient. In the next section it will be demonstrated that, when *population vector coding* is used to combine the responses of the filter bank, only a small number of filters are required in order to achieve an accuracy of orientation estimation considerably higher than that of an individual filter.

3.2 Local Orientation Estimation by Population Vector Decoding

The concept of a population vector was originally introduced by Georgopoulos and colleagues to describe the representation of limb movements in the motor cortex (Georgopoulos et al., 1986). Adapting the authors' definition to a bank of n Gabor filters, the population vector is computed as follows. The filter bank consists of pairs of filters with even symmetry (cosine part) and odd symmetry (sine part), also called *quadrature pairs*. Consider a wavelength λ. Let $G(x, y; \psi_i, \lambda)$ be the response modulus ("energy") of the quadrature pair of Gabor filters of orientation ψ_i. Let $\vec{e}_i = (\cos\psi_i\ , \sin\psi_i)^T$ be the unit vector in the direction ψ_i. Then the population vector $\vec{p}$ is defined as

$$\vec{p}(x, y) = \sum_{i=1}^{n} G(x, y; \psi_i, \lambda)\ \vec{e}_i\ , \tag{3.3}$$

which means that each filter is represented by a two-dimensional component vector. The vector orientation and magnitude are given by the preferred orientation ψ_i and the response magnitude (modulus) $G(x, y; \psi_i)$ of the filter at location (x, y), and the population vector is the sum of the n filter vectors.

However, equation (3.3) cannot directly be applied, since the filters are only sensitive to orientation, rather than direction; i.e., there is a 180°-ambiguity. Consequently, the population vector is computed using the scheme in Figure 3.4. The orientation variable is *de*coded by determining the angle of the population vector θ_{pop}, which is given by:

$$\theta_{pop}(x, y) = \arctan\left(\frac{p_y(x, y)}{p_x(x, y)}\right)\ . \tag{3.4}$$

The magnitude of the population vector, $\|\vec{p}(x, y)\|$, is related to the response "energy" of the filter bank at position (x, y). If evaluated at contour locations, i.e., local maxima of $\|\vec{p}\|$, θ_{pop} gives an estimate of the local tangent angle. Theoretically, the coding

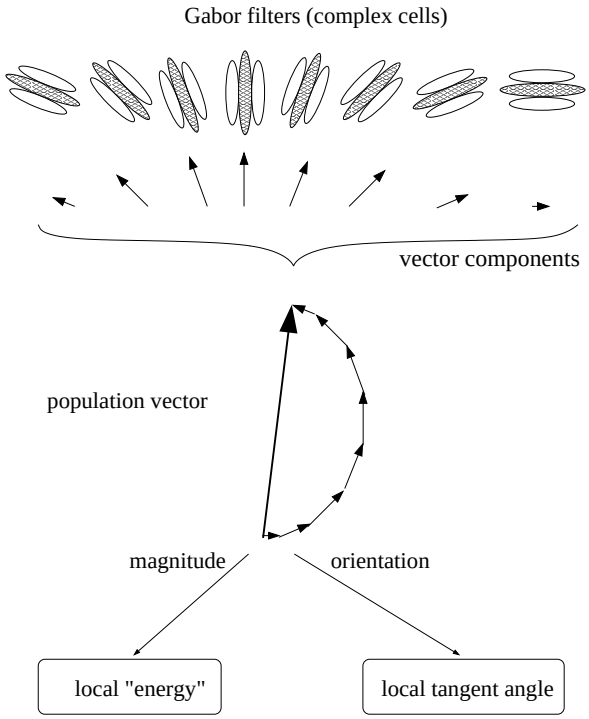


Figure 3.3: General scheme for population coding with a bank of Gabor filters. Each filter is represented by a component vector. The vector orientation and magnitude are given by the preferred orientation and the response magnitude of the filter. The population vector is the vector sum of the components. Its orientation gives an estimate of the local orientation at the considered position. The magnitude characterises the overall response strength of the ensemble.

error can be made as small as desired by applying a larger number of filters. However, computational cost and discretisation errors in digital images limit the optimal number of filters, in practice.

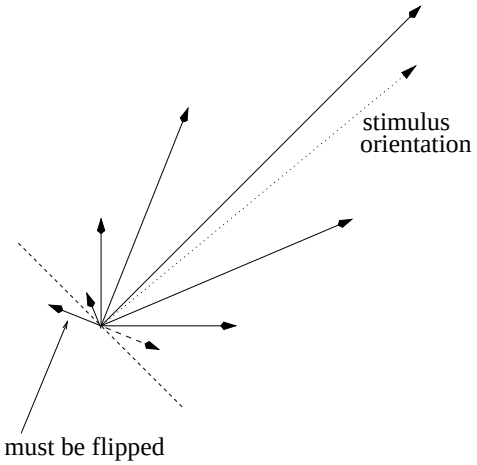


Figure 3.4: Vector coding of orientations. Orientations are restricted to the range $0° - 180°$. Therefore, the vector components are computed with respect to a symmetry axis, in this case the orientation of maximum response. Components outside the $\pm 90°$ range around the axis have to be "flipped" back into that range to enforce a symmetrical arrangement. A component perpendicular to the symmetry axis (i.e., on the dashed line) would effectively cancel itself out and can thus be ignored.

3.3 Matched Filtering and the Extraction of Oriented Energy

Steerable filters have been introduced by Freeman and Adelson (1991) to reduce the computational cost of contour representation. Instead of applying a large number of filters to cover the full range of orientations, only a small set of basis filters is used, from which the response of a filter of arbitrary orientation can be synthesised. The approach is feasible for Gaussian derivatives, but Gabor filters are known to be only *approximately* steerable (Shustorovich, 1994).

In this section, it is shown how the Gabor filter bank can be made steerable, in the sense that the orientational energy can be approximated by means of the magnitude of the population vector. To accomplish this, it is necessary to move beyond the linear vector population coding and to make explicit use of the orientation tuning curve.

Originally, the population vector was used to extract a vectorial quantity, e.g., the position $\vec{x}$ of a limb in three-dimensional space (Georgopoulos et al., 1986). Here the population vector $\vec{p}$ can directly represent the position vector, i.e.,

$$\vec{p} = \vec{x}\,.$$

The population vector and the encoded position vector are expected to be equal in direction *and* magnitude. The response, r_i, of a single direction encoding motor neuron is modelled as a projection of the population vector onto the corresponding unit vector:

$$r_i = \vec{p} \cdot \vec{e}_i\,.$$

The situation with contour representation using Gabor filters is different. There are two separate quantities, orientation and filter response energy, to be merged into a contour representing vector. Its orientation should be equal to the local characteristic orientation of the contour (tangent angle), which is, of course, only well defined if there is unimodal anisotropy in the neighbourhood of the considered location. The degree of anisotropy can, however, be judged from the distribution of filter responses by means of a reliability criterion of the orientation measurement derived in section 3.5.

The magnitude of the contour representing vector shall here be defined as the response of the Gabor filter tuned to the very tangent orientation, θ_c, of the contour, G_{max}. In the literature on steerable filters, this is usually referred to as the "matched filter" (Perona, 1992). The contour vector at a particular point (x, y) is then

$$\vec{c} = G_{max}(x, y; \lambda)\vec{e}_c\,,$$

where $\vec{e}_c$ is the unit vector in the direction of the contour. G_{max} is the orientational energy and therefore a measure of the contour contrast. The question is how the contour vector can be decoded from the filter "population".

The orientation can easily be obtained from equation (3.4). However, as Gabor filters (and visual cortical neurons) have non-cosinusoidal orientation tuning (section 3.1.2, (Swindale, 1998)), the relationship between the magnitudes of the contour vector and the population vector is more complicated than in Georgopoulos' approach (Georgopoulos et al., 1986). The magnitude of the population vector does not equal the orientational energy, as one would expect in linear vector coding, since the filter responses cannot be described as dot-products between the contour vector and the filter component vector. In order to derive the oriented energy, the population vector has to be modelled by means of the tuning curves.

The response of a filter of orientation ψ to an input image with local contour orientation θ_c at (x, y) can be described by G_{max} and the normalised tuning functions f given by equation (3.2):

$$G(x, y; \psi, \lambda) = G_{max}(x, y; \lambda) f(\theta_c; \psi) \,. \tag{3.5}$$

Assuming that the population angle θ_{pop} (the measured orientation) is close to the true contour orientation ($\theta_c \approx \theta_{pop}$), a model population vector $\vec{m}$ is calculated from the *normalised* tuning functions:

$$\vec{m} = \sum_{i=1}^{n} f(\theta_{pop}; \psi_i) \vec{e}_i \,. \tag{3.6}$$

Inserting the response model (3.5) into the definition of the population vector (3.3) and comparing the result with (3.6) yields

$$G_{max} = \frac{|\vec{p}|}{|\vec{m}|} \,. \tag{3.7}$$

Since G_{max} and the tuning properties are known, steerability has been achieved in those image regions where the degree of anisotropy is high enough, i.e., where the response profile of the filter outputs closely resembles the tuning curve. The filter response for any orientation can thus be calculated from equation 3.5.

3.4 Performance of Orientation Estimation

In this section, the accuracy of the orientation estimate is investigated, as well as its dependence on tuning width and noise level. The algorithm is tested on artificial images using filter-banks of different sizes. The test images and the filter wavelength are the same as those previously described in Section 3.1.2. The filter banks consist of 8, 16 and 32 Gabor filters. Figure 3.5 shows the root mean square (rms) error of the population angle $\langle \delta\theta_{pop} \rangle_{rms}$ as a function of the tuning half-width of the applied Gabor filters. The error increases when, given the number of filters, the tuning width is too small to guarantee sufficient filter overlap to cover the entire range of 180 degrees. In the experiments, this limit is not reached with 32 filters. The error seems to continuously decrease further for decreasing tuning width. However, according to the corresponding uncertainty relation, the tuning width decreases at the expense of an increasing receptive field size and, consequently, a lower spatial resolution.

For large tuning widths the envelope parameter is so small that the entire receptive field consists of only a few pixels, and discretisation errors become noticeable.

In conclusion, eight filters can be considered sufficient for practical purposes, since the computational cost is significantly lower and the precision only slightly smaller than with 16 filters.

Compared to the tuning width of a single Gabor filter, the population vector estimate of stimulus orientation is very accurate. The resulting rms-deviation of the angle of the population vector from the ground truth value of the stimulus orientation is only $\langle \delta\theta_{pop} \rangle_{rms} \approx 1^\circ$, which should be compared with the half-width of the tuning curve for the most suitable filter ($w \approx 17^\circ$). The error of the population coded orientation estimate consists of two components: the coding error due to the limited number of filters and the discretisation error caused by the pixelation of digital images. Moreover, the measured rms-error is consistent with simulations by Vogels (1990).

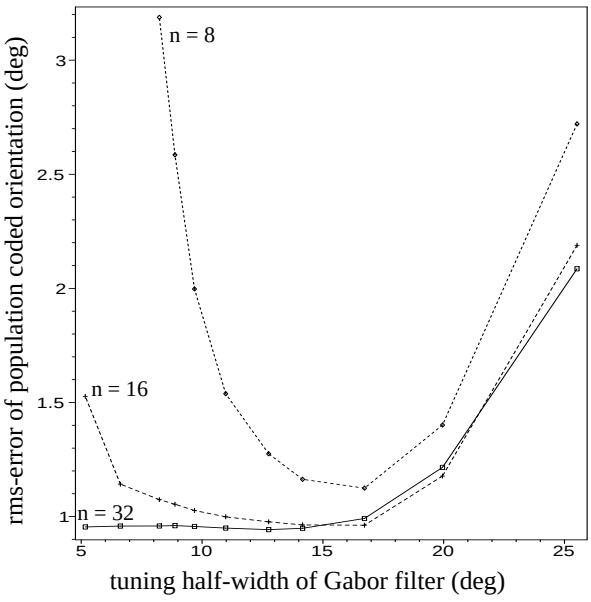


Figure 3.5: Dependence of the rms-error of orientation of the population vector on the tuning half-width for 8, 16 and 32 filters. The wavelength is $\lambda = 8$ pixels. The rms-error reaches a minimum at a tuning width similar to that of cells in striate cortex. Note that the abscissa starts with a tuning half-width of $w = 5^\circ$ rather than $w = 0^\circ$, due to the inverse proportional relation between filter size σ_e and tuning width w, which would require an infinite filter mask to obtain $w = 0^\circ$. The practical lower bound for the tuning width is determined by the upper bound for the mask size, which itself depends on the spatial extent of the structures to be detected in the input image.

3.4.1 Noise robustness

In another experiment, sensitivity to additive Gaussian noise was investigated. Theoretically, there are two essential weaknesses of the population vector method in the presence of noise: the problem of noise contributions from unstimulated units and the potentially small ratio of tuning width to sensor array extent (Snippe, 1996). However, in the context of orientation estimation with a filter bank, neither are of much concern. The range of orientations $[0, \pi]$ (extent of the sensor array) is fixed and can easily be covered by a relatively small number of filters. Moreover, unlike with a linear variable, the circular topology of the sensor array implies that tuning curves located at the ends of the interval $[0, \pi]$ do overlap. Consequently, there are few unresponsive units, unless a small tuning width is chosen, which is disadvantageous even without noise (see section 3.4).

Figure 3.6 shows the rms-error as a function of the noise variance for different numbers of filters (8,16 and 32). The dependence is roughly linear for all three filter banks, with no significant difference in noise sensitivity.

3.4.2 Comparison to Other Methods in the Literature

Buse and associates (Buse et al., 1996) have developed a method for simultaneous measurement of orientation and length of straight lines. Using the real parts (cosine part) of a bank of Gabor filters, they compute an orientation estimate from an interpolation of filter responses at the line end points. Since they use 256×256 grey-scale images of single lines, their data are directly comparable to the results obtained by population coding.

Since Buse and associates do not report the number of filters involved, their results have to be compared with filter banks of different sizes. It seems likely, however, that a good interpolation result requires a number of filters similar to that used for population

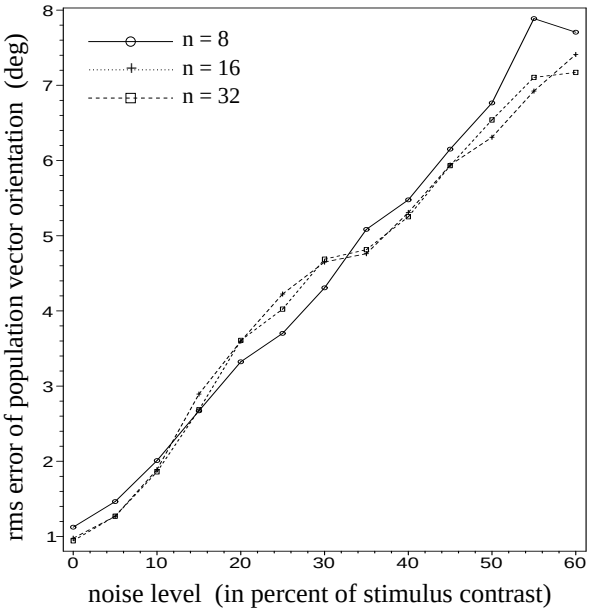


Figure 3.6: The rms-error as a function of the noise variance for additive Gaussian noise using 8, 16 and 32 filters. The method seems to be quite robust. The filter banks show no significant difference in noise sensitivity.

coding. Unfortunately, the authors do not present a detailed statistical analysis of their data. No root mean square error nor a similar measure of the accuracy of the orientation measurement is given, and the limited amount of data given does not allow a reliable calculation of any mean error.

Because of the limited data given by Buse and Caelli, an exact assessment of the accuracy in comparison with population vector coding is difficult. An essential disadvantage of their algorithm is the restriction to straight lines. The population coding method can measure local tangent orientation at any point on a curve, provided that the radius of curvature is larger than the "receptive field" size of the filter. On the other hand, their method also performs a length measurement by means of different filter frequencies.

3.5 The Uncertainty of Orientation Information

The response profile of the filter bank, i.e., the angular distribution of filter outputs at a given point in the input image, contains valuable information of the local contour structure. Zemel and colleagues proposed to represent certainty of local information in terms of the sum of responses (Zemel et al., 1998). In the notation used in this thesis,

$$C(x,y) = \sum_{i=1}^{n} G_i(x,y) \Big/ \left(\sum_{i=1}^{n} G_i \right)_{max}, \tag{3.8}$$

where the denominator is some global maximum of the summed responses. However, this measure only depends on response energy (contour contrast) and cannot discriminate between low contrast contours and intense noise. Also, points of multimodal anisotropy, such as corners (points of high curvature) and junctions, can produce high responses in the filter bank, though local tangent orientation is ill-defined.

It is argued in this thesis that the "sharpness" of the response profile is more suitable to characterise the reliability of the local orientation estimate, as it is contrast independent. Thus contour contrast and certainty are treated as two separate pieces of information. In fact, there is evidence that perceived contrast and the appearance of contours is not so closely linked as is commonly assumed (Hess et al., 1998).

Figure 3.7 shows the response profile of the filter bank at a number of different points in a natural image. Despite the fact that the response profiles are normalised, the quality of the edge (degree of anisotropy), and thus the expected reliability of orientation measurement, is well-reflected in the width of the profile. Accordingly, certainty should be measured in terms of the angular concentration of the response energy around the population vector orientation. At a contour, the response energy of the filter bank can be assumed to be clustered around the contour orientation. Therefore the average of the cosines of orientation differences is used, weighted by the responses:

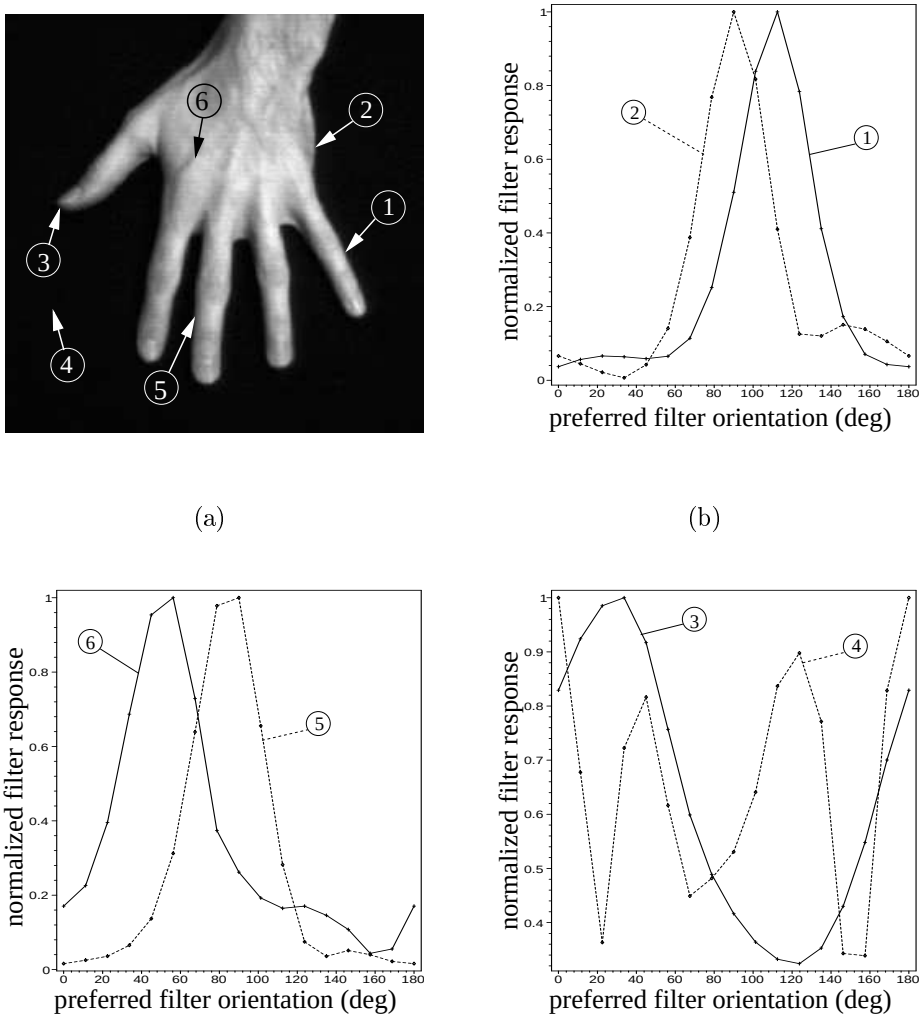


Figure 3.7: Normalised response profiles of the filter bank (16 filters, $\lambda = 8$ pixels, $\sigma_e = 0.6 \times \lambda$) at different points in a natural image **(a)**. Note how the distribution of filter responses reflects the quality of the edge. The width of the response profile allows the distinction of noisy regions from weak contours, independent of their contrast level. **(b)** The half-widths at well pronounced edges (points 1, 2) are very similar to the tuning width of the filter obtained from synthetic line images. Compare $w = 17.2°$ with $w_1 = 19.8°$, $w_2 = 21.5°$. **(c)** Points with less intensity gradient, such as (6) yield a wider profile. Compare $w_6 = 24.2°$ with $w_5 = 21.3°$. **(d)** An even wider profile is obtained in regions with high curvature, such as point (3); $w_3 = 52.9°$. In locations without any orientational structure (point 4), the response energy is spread irregularly over the entire range of orientations.

$$C(x,y) = \frac{\sum_{i=1}^{n} G_i(x,y)\ \cos\left[\frac{\pi}{2} - |\frac{\pi}{2} - \Delta\psi_i|\ \right]}{\sum_{i=1}^{n} G_i(x,y)}$$

$$\Delta\psi_i = |\psi_i - \theta_{pop}|\,, \tag{3.9}$$

where ψ_i is the filter orientation and θ_{pop}, the population coded contour angle. The certainty measure is related to the *resultant length* in directional statistics (Mardia, 1972), a measure of concentration of a set of random samples around its mean direction, defined by the following expectation value:

$$\rho = E\Big\{\cos(\theta - \langle\theta\rangle)\Big\}\,.$$

In this expression, the cosine of the difference between mean direction $\langle\theta\rangle$ and a random sample θ is analogous to the squared difference $(x - \langle x\rangle)^2$. The certainty measure effectively has an upper bound less than unity, since the maximum degree of response clustering is limited by the tuning width. Additionally, every certainty value below 0.5 signals total unreliability, since $C = 0.5$ corresponds to 45°, and any response clustering further than 45° away from the measured orientation θ_{pop} simply indicates that there is a multimodal distribution. The certainty measure could be normalised by dividing it by the largest certainty value detected.

3.6 Tangent Fields

In this section, the population coding techniques for the measurement of orientation, response energy and certainty are combined to obtain tangent fields. The results are promising not only for the purpose of contour detection, but also with respect to the representation of tangent flow fields.

Figure 3.8: An Infra-red aerial image of Luton airport (512×512 pixels).

Figure 3.9 shows a detail from the tangent flow field corresponding to an infra-red aerial image obtained with different methods of orientation measurement. The comparison includes Zucker's method, which uses second Gaussian derivatives, (Zucker et al., 1988), selection of tangent orientation from S-Gabor filters (Leite and Hancock, 1997), and population vector coding.

Note that the population vector approach is able to recover some fine details in the flow. The flow field follows even small structures in the intensity pattern, such as the "vortex" in the upper left, that have been lost by the other algorithms due to smoothing.

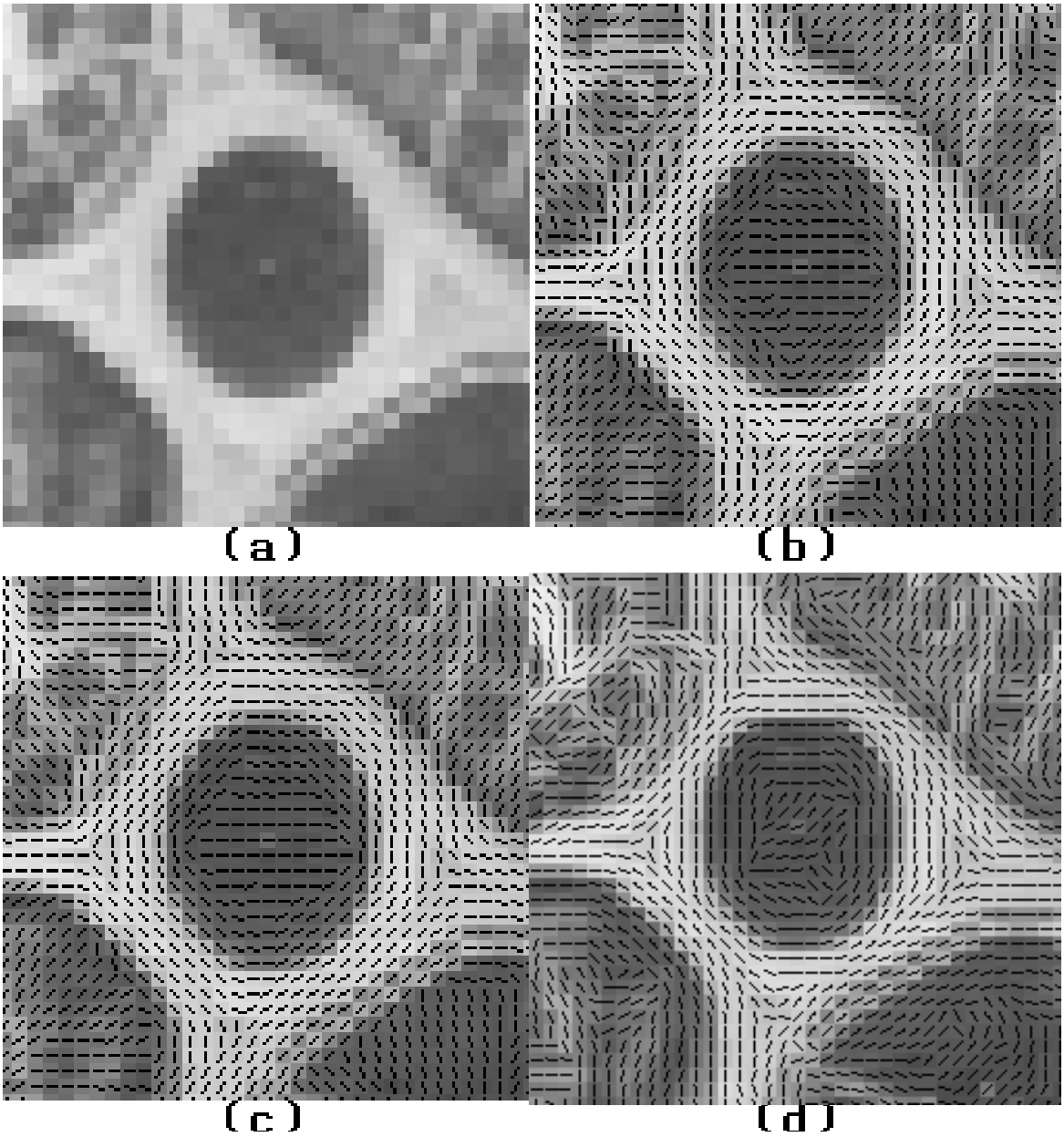


Figure 3.9: Magnified detail (roundabout) of Fig. 3.8. Each grey square represents a pixel. **(a)** original, **(b)** with overlaid tangent field following (Zucker et al., 1988), **(c)** tangent field obtained by selecting the strongest response from 8 S-Gabor filters, **(d)** tangent field from population vector coding with 8 Gabor filters ($\lambda = 3$ pixels). The population vector approach preserves some fine details in the flow field that are lost with the other methods as a consequence of insufficient angular resolution or over-smoothing. Images (a)-(c) after Leite and Hancock (1997).

3.6.1 Contour Representation: Tangent selection in natural images

In order to represent contours by means of local line segments from the tangent field, it is necessary to select those line segments on the crest lines of the response energy landscape given by the magnitude of the population vector. In the terminology of Parent and Zucker, this problem is referred to as the search for *lateral maxima* (Parent and Zucker, 1989).

In a nutshell, the feature selection algorithm used in this chapter first performs the local maximum search on the product of certainty and response energy, $|\vec{p}(x,y)|C(x,y)$, rather than energy $|\vec{p}(x,y)|$ alone. Thus points of high curvature or junctions, where the orientation measurement is not well-defined, are excluded, as well as noisy regions where virtually no orientational structure is present. In a subsequent step, spurious parallels in the contour neighbourhood are eliminated through competition among neighbouring parallel line segments. The remaining points undergo thresholding. As a result, only points of high contrast and high certainty "survive".

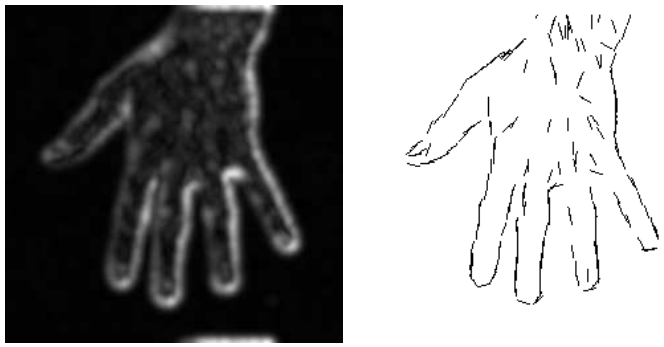
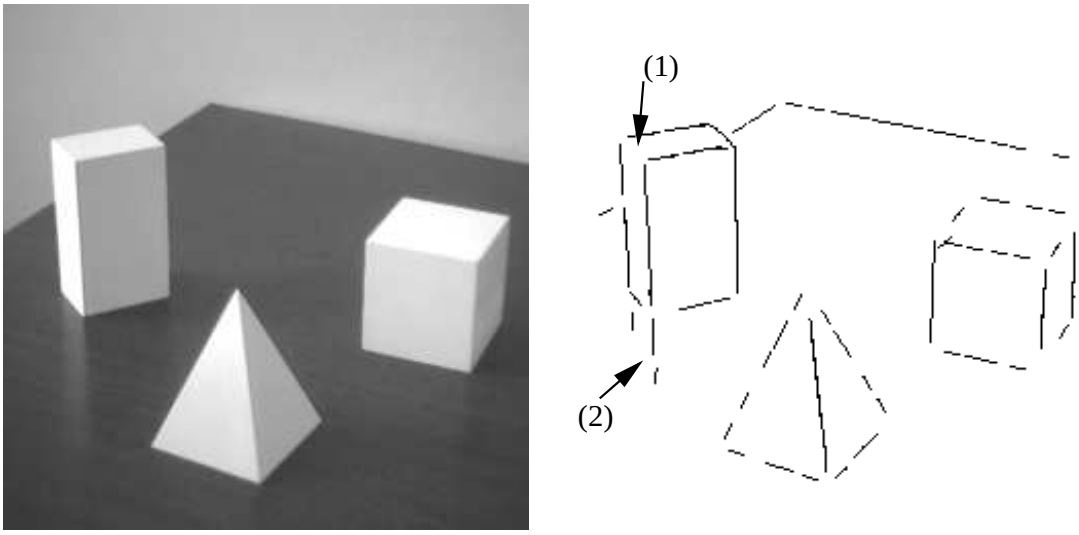


Figure 3.10: Contour representation from population coded tangent fields. **(a)** the orientational energy map. **(b)** the contour tangent field of a natural image (human hand, see Fig. 3.7 a). The tangents in Image (b) represent the local orientation at "key points", i.e., local maxima of the product of directional energy and certainty.

In order to obtain such an initial set of *key points*, a grid is placed on the map of the product $|\vec{p}(x,y)|C(x,y)$. Within each segment a local maximum is determined. Interpreting the map of magnitude $\times$ certainty as a landscape with the edges forming ridges, this is a search for the highest point on the ridge within each grid segment. However, the grid discretisation can produce artifacts if the ridge is just outside the boundary of the considered grid segment. The local maximum will then be only a point on the slope (the real local maximum is located in a neighbouring segment). In order to avoid such false maxima, every candidate must be verified by checking for a local maximum in a new neighbourhood centred around the candidate. If the candidate is still found to be a local maximum, it is accepted as a key point; otherwise, it is erased.

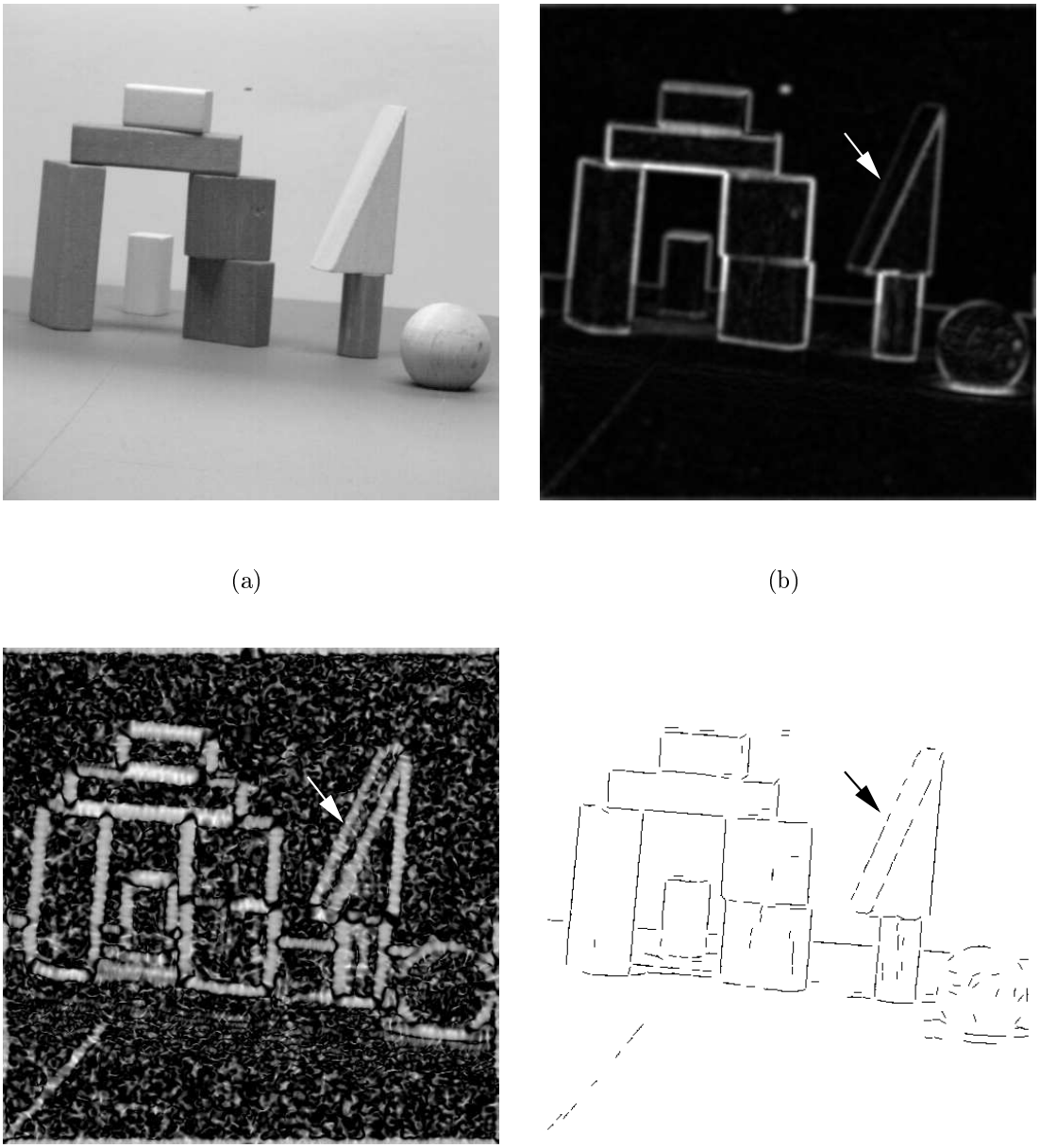
Once these "key points" of reliable contour information and the corresponding tangent angles are determined, they provide a symbolic representation in terms of local line segments (Figure 3.10 and 3.11). Moreover, they could serve as an initialisation of nodes in a graph representation and be further updated by more global constraints.

Figure 3.12 shows each of the different steps in the procedure of contour tangent field extraction. The left contour of the triangular block, denoted by an arrow, appears very faint in the map of the population vector magnitude. However, it is clearly visible in the certainty map, since response profiles along the entire length of the contour tend to cluster tightly around the edge orientation, despite the low intensity gradient. Even though the information about the weak contour is, in principle, present in the energy map (Fig.3.12 b), it is much easier to establish a general threshold for the product of certainty and magnitude than for magnitude alone.



(a) (b)

Figure 3.11: **(a)** An image with several polyhedra (256 ×256 pixels). **(b)** The corresponding contour tangent field extracted with λ=8 pixels. Only extremely faint contours are not represented, such as the short edge (1). Additionally, the reflections on the table surface create some additional line segments. Note that corner points and junctions are left blank.



(a) (b)

(c) (d)

Figure 3.12: **(a)** Another image (512 ×512 pixels). **(b)** The magnitude of the population vector. The weak left contour of the triangular block (arrow) is almost lost. **(c)** The certainty map. As expected for a second order measure, more noise is present. The weak contour is clearly visible, since response profiles all along the contour tend to cluster tightly around the edge orientation, despite the low intensity gradient. **(d)** In the contour tangent field the weak edge has been recovered without amplifying noise.

3.7 Conclusions

Is has been shown that population vector coding with Gabor filters can be used to accurately estimate the orientation of edges and lines in grey-scale images. Moreover, based on an analysis of local filter response profiles, a certainty measure has been defined in terms of the distribution of response energy around the population vector. The certainty measure is, by definition, contrast independent, and therefore provides an additional piece of information that signals the presence of an edge/line. The certainty measure also helps to avoid the calculation of wrong orientation estimates in unsuitable locations, such as corner points and junctions.

An important issue is the biological plausibility of the population vector approach. The fact that a population vector interpretation allows read-out of the information encoded by a neural ensemble through the experimenter (Georgopoulos et al., 1986) does not mean that such decoding is actually performed in the brain (Pouget and Zhang, 97; Oram et al., 1998; Lehky and Sejnowski, 1998). It is more likely that distributed coding is maintained to secure robustness against noise and loss of neurons (Snippe, 1996).

It is *not* claimed that the algorithm models aspects of the cortical processing of orientation information. However, all the operations necessary for computing a population vector could easily be realized by cortical neural networks. Also, it turns out that the optimal performance of orientation estimation by the system is reached when the tuning width of the filters resembles that of striate cortical cells.

Another criticism stems from statistical estimation theory. Due to the non-cosine orientation tuning curves, the population vector is not an *efficient estimator*, i.e., it does not attain the Cramér-Rao bound (Snippe, 1996). This means that the angular measurement is not optimal in the sense that it does not have the lowest (circular) variance and, consequently, minimal root mean square error. The optimal rms-error

would be achieved within a maximum likelihood framework that takes into account the actual shape of the tuning curve (Sanger, 1996; Oram et al., 1998).

A serious limitation is the fact that regions of multiple orientations, such as corner points, cannot be accurately described in terms of a single population vector. In the next chapter, the population coding approach is therefore extended towards a probabilistic interpretation of population coding, in order to recover full probability distributions over local orientation.

Chapter 4

Probabilistic Population Coding

Probabilistic population decoding of sensory input data is concerned with the extraction of a probability density function from the responses of a set/population of filters/neurons tuned to a stimulus property. A major advantage of the probabilistic approach over other methods is that it is not restricted to the decoding of a unique value, since it can represent uncertainty and ambiguity in the encoded variable through a multi-modal distribution.

Substantial theoretical contributions in probabilistic population coding have been made by Sanger (1996) as well as Zemel, Dayan and colleagues (Zemel et al., 1998; Zemel and Pillow, 2000; Zemel and Dayan, 1999; Zemel et al., 1995), which are mainly concerned with modelling biological information processing, particularly the perception of motion and orientation.

In this chapter, probabilistic population coding is applied to the detection and representation of local edge orientation. The ability to represent ambiguous inputs is used to extract multiple orientations in corner points and junctions. Thus it is possible to overcome the restrictions of the population vector approach, which only allows the extraction of tangent fields from unimodal response profiles.

Unlike neurons in biological models of population coding, where firing rates are

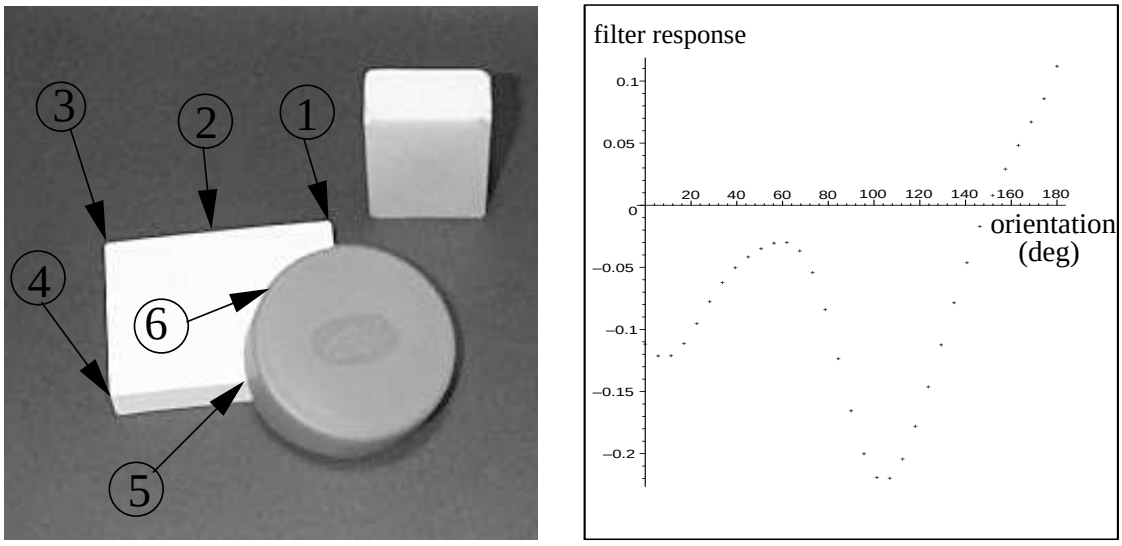
regarded as random variables following Poisson statistics (Zemel et al., 1998), or a related probabilistic model (Itti et al., 1998), the filters are treated as deterministic operators fed with a stochastic visual input. Even though the filtering process is deterministic per se, the responses themselves become stochastic, since they are functions of a random variable: the local contour orientation θ.

In order to recover the whole pdf of the local orientation, $p(\theta)$, representing the uncertainty and ambiguity of θ, the pdf parameters have to be estimated from the given set of filter responses, which 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 can determined so that the given filter responses are most likely. The pdf of the input variable is thus decoded from the population activities (filter outputs).

4.1 Gabor Filter Response Profiles at Locations of Multiple Orientation

The response profile of a bank of Gabor filters extracted at locations of multiple orientation contains several peaks, which are more or less separated, depending on the tuning width of the filters applied. An adequate choice of filter parameters allows the discrimination of at least two response extrema corresponding to the different orientations of the edges meeting or overlapping at the considered point. Figure 4.1 (b) shows the response profile extracted at point (5) in Figure 4.1 (a).

Only odd symmetric Gabor filters ("edge detectors") are considered throughout this chapter, but a similar analysis can be made for the even-symmetric counterpart



(a) (b)

Figure 4.1: **(a)** A real image (256×256 Pixels) with several intensity features, such as edges, corners and T-junctions. **(b)** The response profile of the Gabor filter bank at a T-junction (point 5). The ensemble activities reflect the two principal orientations.

("line detector"). The reason for returning to the components of the quadrature pair, rather than continuing with magnitudes, is that a linear filter is required. If the filters are strictly linear, one can assume that the principle of superposition holds, which implies that the response profile for complex intensity structure where several edges coincide (e.g., at corner points) is a linear combination of the response activities for the individual edge components. This assumption is crucial for the following derivations.

It is certainly valid for Gabor filters with even or odd symmetry but already is violated with Gabor moduli, since the magnitude of a complex filter response is calculated by a Pythagorean sum, which is a non-linear operation.

4.1.1 The tuning function of odd-symmetric Gabor filters

The tuning function of an odd-symmetrical 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. This characteristic spatial frequency of an edge's intensity structure at a particular location influences the tuning width of the filter responses. For moderately curved stimuli, the above argument is still valid within the effective range of the filter mask (i.e., where the Gaussian envelope is significantly above value zero).

Accordingly, the "input image" used to calculate the tuning function is a sinusoid of infinite extent. (In practice, this is to be understood in the sense of periodic continuation applied during the subsequent fast Fourier Transform.) For practical purposes, the corresponding wavelength λ_s (in pixels) will be used, rather than the frequency. The sinusoid's orientation θ is defined in terms of the orientation of its wave vector, the magnitude of which will be denoted by $k_s = 2\pi/\lambda_s$. Hence, the sinusoid is written as

$$S(x, y) = \sin\left[k_s \left(x \cos\theta + y \sin\theta\right)\right] \tag{4.1}$$

For simplicity, and without loss of generality, a Gabor filter is chosen with wavelength λ_f and vertical orientation. Additionally, the quantity $k_f = 2\pi/\lambda_f$ is introduced. Thus

$$\mathcal{G}_{odd}(x, y) = \exp\left(-\frac{x^2}{2\sigma_w^2} - \frac{y^2}{2\sigma_e^2}\right) \sin(k_f x)\,. \tag{4.2}$$

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

$$
\begin{aligned}
f(\theta) &= (\mathcal{G}_{odd} * S)(0,0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{G}_{odd}(x,y)S(x,y,\theta)\,dx\,dy \\
\Longrightarrow f(\theta) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma_w^2} - \frac{y^2}{2\sigma_e^2}\right)\sin(k_f x)\ \sin\left[k_s\left(x\cos\theta + y\sin\theta\right)\right] \quad (4.3)
\end{aligned}
$$

The rather counterintuitive choice of Cartesian, instead of polar, coordinates is necessary, since the filter kernel is not polar separable; consequently, the resultant integrals in the angular and radial variable would be virtually intractable.

Expanding the sinusoid yields:

$$
\sin\left[k_s\left(x\cos\theta + y\sin\theta\right)\right] = \sin(k_s x\cos\theta)\cos(k_s y\sin\theta) + \cos(k_s x\cos\theta)\sin(k_s y\sin\theta)
$$

Hence, separation of integrations in x and y is straightforward:

$$
\begin{aligned}
f(\theta) &= \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma_w^2}\right)\sin(k_f x)\sin[k_s x\cos\theta]\,dx \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma_e^2}\right)\cos[k_s y\sin\theta]\,dy \\
&\quad + \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma_w^2}\right)\sin(k_f x)\cos[k_s x\cos\theta]\,dx \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma_e^2}\right)\sin[k_s y\sin\theta]\,dy
\end{aligned} \quad (4.4)
$$

The second double integral vanishes due to the y-integration of an odd-symmetrical function. In Gradstein and Ryshik (1981), the following integrals (No. 3.898 1. and 3896 4.) are given:

$$\int_{-\infty}^{\infty} e^{-\beta x^2} \sin(ax)\sin(bx)\, dx = \frac{1}{2}\sqrt{\frac{\pi}{\beta}}\left\{\exp\left(-\frac{(a-b)^2}{4\beta}\right) - \exp\left(-\frac{(a+b)^2}{4\beta}\right)\right\}$$

$$\int_{-\infty}^{\infty} e^{-\beta y^2} \cos(by)\, dy = \sqrt{\frac{\pi}{\beta}}\, \exp\left(-\frac{b^2}{4\beta}\right), \qquad \text{for } ab > 0\ ,\ \beta > 0\ .$$

This yields:

$$f(\theta) = \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma_w^2}\right) \sin(k_f x)\sin[k_s x\cos\theta]\, dx \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma_e^2}\right) \cos[k_s y \sin\theta]\, dy$$

$$= \frac{1}{2}\sqrt{2\pi\sigma_w^2}\left\{\exp\left(-\frac{(k_f - k_s\cos\theta)^2 2\sigma_w^2}{4}\right) - \exp\left(-\frac{(k_f + k_s\cos\theta)^2 2\sigma_w^2}{4}\right)\right\}$$
$$\times \sqrt{2\pi\sigma_e^2}\, \exp\left(-\frac{2k_s^2\sin^2\theta\,\sigma_e^2}{4}\right)$$

$$= \pi\sigma_w\sigma_e\left\{\exp\left(-\frac{\sigma_w^2}{2}[k_f - k_s\cos\theta]^2 - \frac{\sigma_e^2}{2}[k_s\sin\theta]^2\right)\right.$$

$$\left. - \exp\left(-\frac{\sigma_w^2}{2}[k_f + k_s\cos\theta]^2 - \frac{\sigma_e^2}{2}[k_s\sin\theta]^2\right)\right\}$$

If the aspect ratio of the filter is set to one, i.e., $\sigma_e = \sigma_w$, this simplifies to:

$$f(\theta) = \pi\sigma_e^2\left\{\exp\left(-\frac{k_f^2\sigma_e^2}{2} + k_f k_s \sigma_e^2\cos\theta - \frac{k_s^2\sigma_e^2}{2}\right)\right.$$
$$\left. - \exp\left(-\frac{k_f^2\sigma_e^2}{2} - k_f k_s\sigma_e^2\cos\theta - \frac{k_s^2\sigma_e^2}{2}\right)\right\}$$

$$= \pi\sigma_e^2 \exp\left(-\left[\frac{k_f^2\sigma_e^2}{2} + \frac{k_s^2\sigma_e^2}{2}\right]\right)\left\{\exp\left(k_f k_s\sigma_e^2\cos\theta\right) - \exp\left(-k_f k_s\sigma_e^2\cos\theta\right)\right\}$$

$$= \underbrace{2\pi\sigma_e^2 \exp\left(-\frac{\sigma_e^2}{2}\left[k_f^2 + k_s^2\right]\right)}_{\hat{f}}\ \sinh\left(\underbrace{k_f k_s\sigma_e^2}_{\kappa_0}\cos\theta\right)$$

$$= \hat{f}\ \sinh[\kappa_0\cos\theta]$$

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 (see Fig.4.2). κ_0 depends on the two known filter properties $k_f = 2\pi/\lambda_f$ and σ_e 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 the following derivations it will be assumed that the tuning functions of all filters are normalised and identical, apart from an angular shift, ψ, indicating their preferred orientation. Hence:

$$f(\theta;\, \psi) = C \sinh[\kappa_0 \cos(\theta - \psi)]\,, \tag{4.5}$$

where $C = 1/\sinh(\kappa_0)$ is a normalisation constant, so that $f(\theta;\, \psi) \in [-1, 1]$. While C is the normalisation constant for the theoretical tuning curve (4.5), the actual filter responses have to be be scaled by $C/\hat{f}$. Let n be the number of filters.

$$r_\nu \mapsto r_\nu \frac{C}{\hat{f}}\,, \qquad \text{where} \quad \nu = \{1, \ldots, n\}$$

The orientation tuning functions of cortical neurons have been shown to be of von Mises type (Swindale, 1998), i.e., of the form

$$f(\theta) \propto e^{\kappa \cos(\theta - \psi)}\,.$$

This is consistent with eqn.(4.5), in the sense that for realistic tuning widths (where κ is not too small)

$$e^{\kappa \cos(\theta-\psi)} \approx e^{\kappa \cos(\theta-\psi)} - e^{-\kappa \cos(\theta-\psi)} = 2 \sinh(\kappa \cos(\theta - \psi))\,, \qquad \text{for} -\tfrac{\pi}{2} < \theta - \psi < \tfrac{\pi}{2}\,,$$

and considering the fact that, unlike Gabor filter responses, cortical neural firing rates are, of course, positive. (In a real neural network the negative part of the tuning

function would be signalled by means of a second neuron sensitive to opposite contrast sign.) Figure 4.3 illustrates the similarity of both functions.

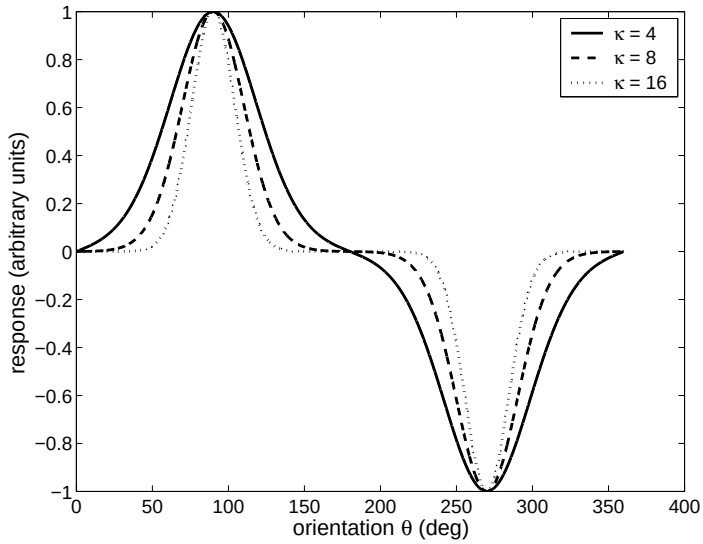


Figure 4.2: Orientation tuning function of an odd-symmetric Gabor filter for three different values of the concentration parameter. The larger κ is, the smaller the tuning width.

4.2 Theory of Probabilistic Population Coding with Gabor Filters

The Gabor filter bank is an ensemble of orientation sensitive units resembling the population of simple cells in a hypercolumn of primary visual cortex. The view of this thesis is that such an ensemble of orientation sensitive units can capture one- as well as two-dimensional intensity features, rendering specific corner or junction detectors unnecessary. A probabilistic population coding framework can, therefore, constitute a generalised edge detector by decoding a potentially multi-modal probability density of local orientation.

Based on the principle of linear superposition (discussed in section 4.1), the underlying probability density of local edge orientation can be modelled by a mixture

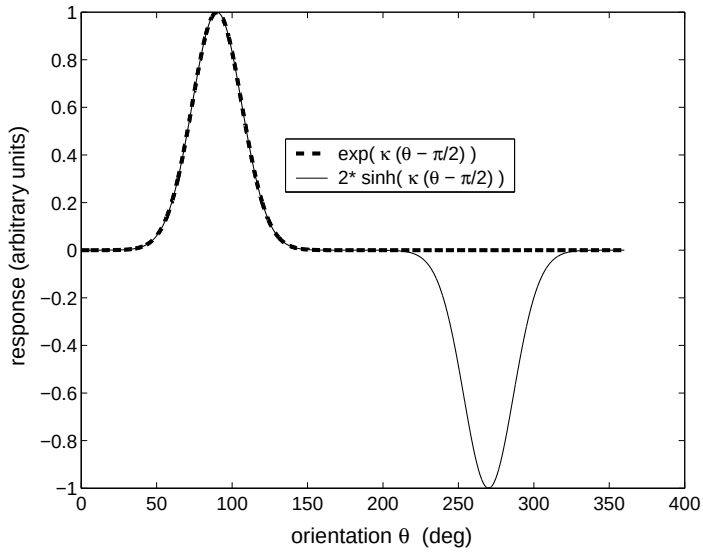


Figure 4.3: Orientation tuning function of an odd-symmetric Gabor filter ($\kappa = 12, \psi = \frac{\pi}{2}$) compared with the corresponding von Mises function which fits many neural orientation tuning curves. Within the range $\theta \in [\psi - \frac{\pi}{2}, \psi + \frac{\pi}{2}]$ the difference is negligible.

distribution. In the case of non-linear filters, a much more complicated model, in terms of a joint density of orientations, would have to be employed. Moreover, the tuning curve would have a more complicated structure wherein the absolute contrast of stimuli would play a crucial role.

4.2.1 Expectation values of filter responses

Let $f(\theta)$ be the orientation tuning function, and let ψ be the preferred orientation of a filter. Assuming superposition holds, the expectation value of the response profile for an arbitrary pdf of the stimulus orientation, $p(\theta)$, is given by (Zemel et al., 1998):

$$\bar{r}(\psi) = \int_0^{2\pi} f(\theta - \psi) p(\theta)\, d\theta\,. \tag{4.6}$$

For symmetrical tuning curves, i.e., $f(\theta - \psi) = f(\psi - \theta)$, the integral in (4.6) can be rewritten as the convolution of tuning function and orientation pdf:

$$\bar{r}(\psi) = \int_0^{2\pi} f(\psi - \theta) p(\theta)\, d\theta\,. \tag{4.7}$$

These equations describe the process of encoding the pdf $p(\theta)$ in the expected response function $\bar{r}$, which is continuous in ψ. The set of average ensemble activities (the average response profile), denoted by $\{\bar{r}(\psi_\nu)\}$, is 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* (Wilson and Lüdtke, 2000), which is, in general, an ill-posed problem (Press et al., 1989) likely to require some kind of regularisation. To account for this, a smoothness prior is used by Zemel et al. (1998). In the following sections, a parametric model of the expected filter responses will be derived based on the orientation tuning function and a mixture model of $p(\theta)$. Within this framework, regularisation is achieved in an *implicit* fashion as a consequence of low model complexity.

4.2.2 A mixture model of local orientation

The encoding equation (4.7) is general, in the sense that it does not restrict $p(\theta)$ to unimodality. Based on the assumption of superposition, the probability density of the stimulus orientation is modelled as a mixture of von Mises distributions (Mardia, 1972), in order to allow accurate representation of ambiguous local orientation (corners, junctions, etc.):

$$p(\theta) = \frac{1}{2\pi} \sum_{i=1}^{m} \frac{P(i)}{I_0(\kappa_i)} e^{\kappa_i \cos(\theta - \bar{\theta}_i)}, \qquad \text{with } \sum_{i=1}^{m} P(i) = 1\,. \tag{4.8}$$

Here I_0 is the modified Bessel function of the first kind and order zero, and the term $1/2\pi I_0(\kappa_i)$ serves as a normalisation factor of the i-th mixture component. Eqn.(4.8) can be considered a circular analogue of the Gaussian mixture density, where the κ_i correspond to the $1/\sigma_i$, the $\bar{\theta}_i$ to the μ_i, and the $P(i)$ are the mixing coefficients.

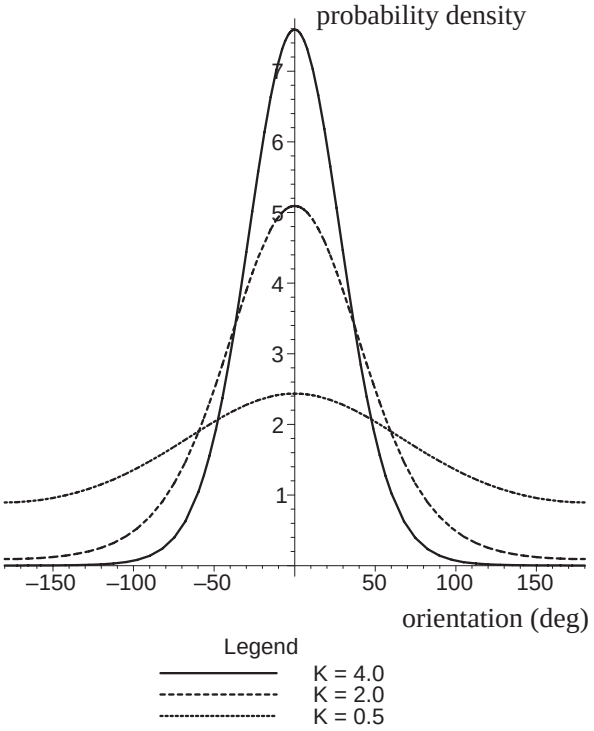


Figure 4.4: Plot of the von Mises distribution.

Figure 4.4 shows a plot of the von Mises distribution for different concentration parameters. The number of mixture components, m, will be limited to two or at most three, describing all essential cases of multiple edge orientation.

Inserting equations (4.5) and (4.8) in (4.7) 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):

$$\bar{r}(\psi) = \frac{C}{4\pi} \sum_{i=1}^{m} \left(\frac{P(i)}{I_0(\kappa_i)} \int_0^{2\pi} \left[e^{\kappa_0 \cos(\theta-\psi)} - e^{-\kappa_0 \cos(\theta-\psi)} \right] e^{\kappa_i \cos(\theta-\bar{\theta}_i)} \, d\theta \right) . \tag{4.9}$$

Therefore, it is necessary to solve integrals of the type

$$\int_0^{2\pi} e^{\pm\kappa_0 \cos(\theta-\psi)+\kappa_i \cos(\theta-\bar{\theta}_i)} \, d\theta \, . \tag{4.10}$$

By means of the substitution $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$ and after

rearranging terms, the exponent with positive sign in the integral becomes:

$$\kappa_0 \cos(\theta - \psi) + \kappa_i \cos(\theta - \bar{\theta}_i)$$

$$= \kappa_0 \cos(\theta)\cos(\psi) + \kappa_0 \sin(\theta)\sin(\psi) + \kappa_i \cos(\theta)\cos(\bar{\theta}_i) + \kappa_i \sin(\theta)\sin(\bar{\theta}_i)$$

$$= \cos(\theta) \underbrace{\left[\kappa_0 \cos(\psi) + \kappa_i \cos(\bar{\theta}_i)\right]}_{=:\alpha_i} + \sin(\theta) \underbrace{\left[\kappa_0 \sin(\psi) + \kappa_i \sin(\bar{\theta}_i)\right]}_{=:\beta_i} \tag{4.11}$$

Analogously, let for $-\kappa_0$:

$$\tilde{\alpha} := -\kappa_0 \cos(\psi) + \kappa_i \cos(\bar{\theta}_i) \quad \text{and} \quad \tilde{\beta} := -\kappa_0 \sin(\psi) + \kappa_i \sin(\bar{\theta}_i)$$

This yields:

$$\bar{r}(\psi) = \frac{C}{4\pi} \sum_{i=1}^{m} \left(\frac{P(i)}{I_0(\kappa_i)} \int_0^{2\pi} e^{\alpha_i \cos(\theta) + \beta_i \sin(\theta)} - e^{\tilde{\alpha_i} \cos(\theta) + \tilde{\beta_i} \sin(\theta)} \, d\theta \right) . \tag{4.12}$$

Using the more general integral No. 3.937 in Gradstein and Ryshik (1981), it follows that

$$\int_0^{2\pi} e^{p \cos x + q \sin x} \, dx = 2\pi I_0(\sqrt{p^2 + q^2}), \tag{4.13}$$

where I_0 is again the modified Bessel function of first kind and order zero. Before applying this result to solve the integrals in (4.12), the terms $\alpha_i^2 + \beta_i^2$ and ${\tilde{\alpha_i}}^2 + {\tilde{\beta_i}}^2$ need to be simplified. With (4.11) one obtains:

$$\begin{aligned} \alpha_i^2 + \beta_i^2 &= \left(\kappa_0 \cos(\psi) + \kappa_i \cos(\bar{\theta}_i)\right)^2 + \left(\kappa_0 \sin(\psi) + \kappa_i \sin(\bar{\theta}_i)\right)^2 \\ &= \kappa_0^2 + \kappa_i^2 + 2\kappa_0\kappa_i \left[\cos(\psi)\cos(\bar{\theta}_i) + \sin(\psi)\sin(\bar{\theta}_i)\right] \\ &= \kappa_0^2 + \kappa_i^2 + 2\kappa_0\kappa_i \cos(\psi - \bar{\theta}_i) \, . \end{aligned} \tag{4.14}$$

Analogously, for $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ one obtains a similar expression (except for the negative sign):

$$\tilde{\alpha_i}^2 + \tilde{\beta_i}^2 = \kappa_0^2 + \kappa_i^2 - 2\kappa_0\kappa_i \cos(\psi - \bar{\theta}_i)\,. \tag{4.15}$$

Next, substituting (4.14) and (4.15), respectively, in (4.13) and inserting the results in (4.12) yields the average response profile:

$$\begin{aligned}\bar{r}(\psi) \;=\; \frac{C}{2}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\,\Bigg\{ & I_0\Bigg(\sqrt{\kappa_0^2 + \kappa_i^2 + 2\kappa_0\kappa_i\cos(\psi - \bar{\theta}_i)}\Bigg) \\ & - I_0\Bigg(\sqrt{\kappa_0^2 + \kappa_i^2 - 2\kappa_0\kappa_i\cos(\psi - \bar{\theta}_i)}\Bigg)\Bigg\}\end{aligned} \tag{4.16}$$

4.2.3 The likelihood of the response profile

In the previous section, the expected response profile, $\bar{r}(\psi)$, has been derived as an expectation value of the tuning function, given a particular set of model parameters of the underlying pdf of the input variable θ:

$$\bar{r}(\psi) = \int_0^{2\pi} f(\theta - \psi)\, p(\theta)\; d\theta\,. \tag{4.17}$$

Alternatively, the expected response profile can also be thought of as the result of averaging the filter responses. 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

$$\bar{r}(\psi) = \int_{-1}^{1} r\, p(r;\psi)\, dr\,. \tag{4.18}$$

The filter responses r_ν, obtained at a particular location in an image, are instances of stochastic variables, even though the filtering per se is a deterministic operation. The randomness of the responses is created solely by the stochastic nature of the input variable, which is fundamentally different from standard biological models of

population coding, where neural firing rates are usually random variables following a Poisson statistics (Zemel et al., 1998) or a related probabilistic model (Itti et al., 1998).

Maximum likelihood estimation

One method of estimating the parameter0s of the mixture model $p(\theta)$ is through a maximum likelihood estimation. Let

$$\Theta = \left\{P(1), \ldots, P(m), \bar{\theta}_1, \ldots, \bar{\theta}_m, \kappa_1, \ldots, \kappa_m\right\}$$

denote a set of mixture parameters. Then, the optimal parameters Θ_{opt} are given by

$$\Theta_{\text{opt}} = \arg\max_{\Theta} \mathcal{L}\{r_1 \ldots r_n | \Theta, \kappa_0\} .$$

Therefore, 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 be able to calculate the total likelihood of a given response profile. Regarding the responses as independent stochastic variables yields:

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

The likelihood function, $\mathcal{L}$, depends on the parameters of the mixture pdf $p(\theta)$, together with the parameter κ_0 specifying the tuning function. The maximum likelihood estimation for the mixture parameters can be performed using standard techniques, such as the EM algorithm. The remaining tuning parameter κ_0 can be obtained:

a) from an initial measurement on a straight edge,

b) together with the mixture parameters in the same MLE process.

Method b) could have interesting implications in a situation where lateral interactions between neighbouring filters are present. In such a situation, the context dependent

change of the tuning width could be observed, i.e., the *effective tuning width*. In this thesis, however, κ_0 will be determined according to a).

4.2.4 The probability distribution of responses

The task now is to find the pdf of a *function of a random variable*, since the filters transform the stimulus orientation, θ, into responses via their tuning function, given by (4.5). According to the general theorem for the probability density of a function of a random variable (Papoulis, 2002, p.130), the pdf of the response profile (preferred orientation ψ) is given by

$$p(r;\psi) = \frac{p(\theta_1)}{|f'(\theta_1;\psi)|} + \frac{p(\theta_2)}{|f'(\theta_2;\psi)|}\,. \tag{4.19}$$

Here 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 $[\psi - \frac{\pi}{2}, \psi + \frac{\pi}{2}]$, given by the inverse of the tuning function. Thus $\theta_{1,2}$ are the solutions of the equation

$$r = f(\theta_{1,2};\psi)\,, \qquad \theta_{1,2} \in [\psi - \tfrac{\pi}{2}, \psi + \tfrac{\pi}{2}]\,. \tag{4.20}$$

Inverting the tuning function, eqn.(4.5), yields:

$$\begin{aligned}
r &= C\sinh[\kappa_0\cos(\theta-\psi)] \\
\frac{r}{C} &= \sinh[\kappa_0\cos(\theta-\psi)] \\
\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right) &= \cos(\theta-\psi) \\
\arccos\left\{\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\right\} &= \pm(\theta-\psi) \\
\Longrightarrow\ \theta_{1,2}(r;\psi) &= \pm\arccos\left\{\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\right\} + \psi\,.
\end{aligned} \tag{4.21}$$

The derivative of the tuning function is given by:

$$f^{'}(\theta;\psi) = \frac{df(\theta;\psi)}{d\theta} = -C\kappa_0 \sin(\theta-\psi)\ \cosh[\kappa_0 \cos(\theta-\psi)]\,. \tag{4.22}$$

Inserting (4.21) into the angular pdf (4.8) yields:

$$p(\theta_{1,2}) = \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\exp\left[\kappa_i \cos\left(\pm\arccos\left\{\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\right\}+\psi-\bar{\theta}_i\right)\right]\,. \tag{4.23}$$

Applying the addition theorem of the cosine and using $\sin(\arccos x) = \pm\sqrt{1-x^2}$, one obtains:

$$\begin{aligned} p(\theta_{1,2}) \;=\; & \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\exp\left[\frac{\kappa_i}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\cos(\psi-\bar{\theta}_i)\right. \\ & \left.\pm\kappa_i\sqrt{1-\left[\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\right]^2}\sin(\psi-\bar{\theta}_i)\right]\,. \end{aligned} \tag{4.24}$$

The second ingredient of theorem (4.19), i.e., $|f^{'}(\theta_{1,2};\psi)|$, is obtained by inserting (4.21) into (4.22):

$$\begin{aligned} |f^{'}(\theta_{1,2};\psi)| \;=\; & \left|-C\kappa_0\ \sin\left(\pm\arccos\left\{\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\right\}+\psi-\ \psi\right)\ \times\right. \\ & \left.\cosh\left[\kappa_0\cos\left(\pm\arccos\left\{\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\right\}+\psi-\ \psi\right)\right]\right|\,. \end{aligned}$$

Using $\sin(\arccos x) = \pm\sqrt{1-x^2}$ again and $\cosh(\sinh^{-1}x) = \sqrt{1+x^2}$, this simplifies to:

$$|f^{'}(\theta_{1,2};\psi)| = C\kappa_0\ \sqrt{1+\left(\frac{r}{C}\right)^2}\ \sqrt{1-\left[\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right)\right]^2}\,. \tag{4.25}$$

Thus using (4.24) and (4.25) along with $e^x + e^{-x} = 2\cosh x$, the pdf of the response $p(r;\psi)$ can be obtained according to theorem (4.19):

$$p(r;\psi) = \frac{1}{\pi C \kappa_0} \sum_{i=1}^{m} \frac{P(i)}{I_0(\kappa_i)} \frac{\exp\left[\frac{\kappa_i}{\kappa_0} \sinh^{-1}\left(\frac{r}{C}\right) \cos(\psi - \bar{\theta}_i)\right]}{\sqrt{1 + \left(\frac{r}{C}\right)^2} \sqrt{1 - \left[\frac{1}{\kappa_0} \sinh^{-1}\left(\frac{r}{C}\right)\right]^2}} \times$$

$$\times \cosh\left(\kappa_i \sqrt{1 - \left[\frac{1}{\kappa_0} \sinh^{-1}\left(\frac{r}{C}\right)\right]^2} \sin(\psi - \bar{\theta}_i)\right) \ . \quad (4.26)$$

This density is, again, a mixture model. However, the mixture components are highly non-Gaussian, making it difficult, though not impossible, to find the parameters. The actual range of the normalised responses is the *open* interval (-1,1), as the density has singularities in $r = \pm 1$.

It is important to note that, due to the structure of $p(r;\psi)$, the *most likely* response profile for a given set of parameters differs from the *average* profile $\{\bar{r}(\psi_\nu)\}$ given by eqn. (4.16). The two would only equal if the responses fluctuated around their mean values following Gaussian distributions. However, the mere fact that response values are bounded, to be more precise $r \in (-1, 1)$, implies that the response pdfs cannot be Gaussian. For this reason, maximum likelihood estimation is the appropriate method of parameter estimation, rather than fitting the response data with the function $\bar{r}(\psi)$ and minimising the squared error,

$$E = \sum_{\nu=1}^{n} \left[r_\nu - \bar{r}(\psi_\nu)\right]^2 .$$

Proof of Normalisation

Though theorem (4.19) implies that the result is a pdf, and therefore normalised, it shall be proved that eqn.(4.26) is, indeed, normalised, i.e.,

$$\int_{-1}^{1} p(r;\psi)\, dr = 1 \, . \quad (4.27)$$

Since $p(\pm 1;\psi) = \infty$, the integral is to be understood in the sense of the limit

$$\int_{-1}^{1} p(r;\psi)\,dr = \lim_{\epsilon\to 0} \int_{-1+\epsilon}^{1-\epsilon} p(r;\psi)\,dr = 1\,.$$

In order to show that (4.27) is satisfied, the following substitution is introduced:

$$\xi(r) := \frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r}{C}\right) \implies d\xi = \frac{dr}{C\kappa_0\sqrt{1+\left(\frac{r}{C}\right)^2}}\,.$$

With $C = 1/\sinh(\kappa_0)$ the limits of the integration then become:

$$\xi_1 \;=\; \xi(-1) = \frac{1}{\kappa_0}\sinh^{-1}\left(\frac{-1}{C}\right) = \frac{1}{\kappa_0}\sinh^{-1}\left(-\sinh(\kappa_0)\right) = -1$$

and, consequently,

$$\xi_2 \;=\; \xi(+1) = 1\,.$$

Thus the integral simplifies to:

$$\int_{-1}^{1} p(r;\psi)\,dr = \frac{1}{\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\int_{-1}^{1}\frac{e^{\kappa_i\xi\cos(\psi-\bar{\theta}_i)}}{\sqrt{1-\xi^2}}\cosh\left(\kappa_i\sin(\psi-\bar{\theta}_i)\sqrt{1-\xi^2}\right)d\xi\,,$$

which can be further simplified by means of the trigonometric substitution $\xi =: \sin\eta$, $d\xi = \cos\eta\,d\eta$, eventually resolving the problem of singular boundaries. Thus the limits of integration are:

$$\eta_1 \;=\; \eta(-1) = \arcsin(-1) = -\pi/2\,,$$

$$\eta_2 \;=\; \eta(1) = \arcsin(1) = \pi/2\,.$$

and the integral becomes:

$$
\begin{aligned}
\int_{-1}^{1} p(r;\psi)\,dr &= \frac{1}{\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\int_{-\pi/2}^{\pi/2}\frac{e^{\kappa_i \sin\eta\cos(\psi-\bar{\theta}_i)}}{\cos\eta}\cosh\left(\kappa_i\sin(\psi-\bar{\theta}_i)\cos\eta\right)\cos\eta\,d\eta \\
&= \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\int_{-\pi/2}^{\pi/2} e^{\kappa_i \sin\eta\cos(\psi-\bar{\theta}_i)}\left(e^{\kappa_i\sin(\psi-\bar{\theta}_i)\cos\eta}+e^{-\kappa_i\sin(\psi-\bar{\theta}_i)\cos\eta}\right)\,d\eta \\
&= \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\left(\int_{-\pi/2}^{\pi/2} e^{\kappa_i[\sin\eta\cos(\psi-\bar{\theta}_i)+\cos\eta\sin(\psi-\bar{\theta}_i)]}\,d\eta\right. \\
&\qquad\left. +\int_{-\pi/2}^{\pi/2} e^{\kappa_i[\sin\eta\cos(\psi-\bar{\theta}_i)-\cos\eta\sin(\psi-\bar{\theta}_i)]}\,d\eta\right) \\
&= \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\left(\int_{-\pi/2}^{\pi/2} e^{\kappa_i\sin[\eta+\psi-\bar{\theta}_i]}\,d\eta+\int_{-\pi/2}^{\pi/2} e^{\kappa_i\sin[\eta-(\psi-\bar{\theta}_i)]}\,d\eta\right) \\
&= \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\left(\int_{0}^{\pi} e^{\kappa_i\cos[\eta+\psi-\bar{\theta}_i]}\,d\eta+\int_{0}^{\pi} e^{\kappa_i\cos[\eta-(\psi-\bar{\theta}_i)]}\,d\eta\right) \\
&= \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\left(\int_{0}^{\pi} e^{\kappa_i\cos[\eta+\psi-\bar{\theta}_i]}\,d\eta+\int_{-\pi}^{0} e^{\kappa_i\cos[\eta+(\psi-\bar{\theta}_i)]}\,d\eta\right) \\
&= \frac{1}{2\pi}\sum_{i=1}^{m}\frac{P(i)}{I_0(\kappa_i)}\underbrace{\int_{-\pi}^{\pi} e^{\kappa_i\cos[\eta+\psi-\bar{\theta}_i]}\,d\eta}_{=2\pi I_0(\kappa_i)} \\
&= \sum_{i=1}^{m} P(i) = 1
\end{aligned}
$$

□

Aside from the phase shift $\psi-\bar{\theta}_i$, which can be neglected due to 2π-periodicity, the integral in the second to the last step is, basically, the integral definition of the modified Bessel function of the first kind and order zero (Abramowitz and Stegun, 1970).

4.2.5 Pdf parameter estimation via the EM-algorithm

Having derived the pdf of responses from the pdf of the local orientation, it is possible to obtain the model parameters by means of a maximum likelihood estimation using the EM-algorithm (Dempster et al., 1977). Due to the complexity of the response pdf, however, the log-likelihood does not lead to significant simplification, as in the Gaussian case. Instead, the update equations are transcendental and require a numerical procedure to solve them for the new parameter values.

Adopting the notation used in Bishop (1995, p.65), the quantity to minimise is

$$\tilde{Q} = -\sum_{\nu=1}^{n}\sum_{i=1}^{m} P^{old}(i|r_\nu)\ln\left\{P^{new}(i)\,p^{new}(r_\nu|i)\right\}\,. \tag{4.28}$$

Here $p^{new}(r_\nu|i)$ is the i-th mixture component of (4.26) for $\psi = \psi_\nu$. According to Bayes theorem, the "old" posterior probabilities $P^{old}(i|r_\nu)$, describing the label assignment of data points to individual mixture components, are given by (Bishop, 1995):

$$P^{old}(i|r_\nu) = \frac{p^{old}(r_\nu|i)P^{old}(i)}{p^{old}(r_\nu)}\,, \tag{4.29}$$

where $p^{old}(r_\nu) = p^{old}(r;\psi_\nu)$ is given by (4.26). Equation (4.29) is the expectation step.

In the maximisation step, the update equations for the mixture parameters are obtained by differentiating (4.28) 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, as described in Bishop (1995), leading to

$$P^{new}(i) = \frac{1}{n}\sum_{\nu=1}^{n} P^{old}(i|r_\nu)\,. \tag{4.30}$$

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},

$$\frac{\partial \tilde{Q}}{\partial \Theta_i^{new}} = -\sum_{\nu=1}^{n}\sum_{k=1}^{m} P^{old}(k|r_\nu)\frac{\partial}{\partial \Theta_i^{new}} \ln p^{new}(r_\nu|k)\,. \tag{4.31}$$

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:

$$-\sum_{\nu=1}^{n} P^{old}(i|r_\nu)\frac{\partial}{\partial \Theta_i^{new}} \ln p^{new}(r_\nu|i) = 0\,. \tag{4.32}$$

According to (4.26), the logarithm of the i-th mixture component for response r_ν is:

$$\begin{aligned}
\ln p^{new}(r_\nu|i) \;=\; & -\ln\left\{\pi C \kappa_0 I_0(\kappa_i^{new})\right\} - \tfrac{1}{2}\ln\left\{1 - \left[\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r_\nu}{C}\right)\right]^2\right\} \\
& + \frac{\kappa_i^{new}}{\kappa_0}\cos(\psi_\nu - \bar{\theta}_i^{\,new})\sinh^{-1}\left\{\frac{r_\nu}{C}\right\} - \tfrac{1}{2}\ln\left\{1 + \left(\frac{r_\nu}{C}\right)^2\right\} \\
& + \ln\cosh\left(\kappa_i^{new}\sqrt{1 - \left[\frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r_\nu}{C}\right)\right]^2}\,\sin(\psi_\nu - \bar{\theta}_i^{\,new})\right)\,.
\end{aligned} \tag{4.33}$$

With the abbreviation $\xi_\nu = \frac{1}{\kappa_0}\sinh^{-1}\left(\frac{r_\nu}{C}\right)$ for the frequently occurring "data term", eqn.(4.33) becomes:

$$\begin{aligned}
\ln p^{new}(r_\nu|i) \;=\; & -\ln\left\{\pi C \kappa_0 I_0(\kappa_i^{new})\right\} - \tfrac{1}{2}\ln\{1 - \xi_\nu^2\} \\
& + \kappa_i^{new}\,\xi_\nu\cos(\psi_\nu - \bar{\theta}_i^{\,new}) - \tfrac{1}{2}\ln\left\{1 + \left(\tfrac{r_\nu}{C}\right)^2\right\} \\
& + \ln\cosh\left(\kappa_i^{new}\sqrt{1 - \xi_\nu^2}\;\sin(\psi_\nu - \bar{\theta}_i^{\,new})\right)\,.
\end{aligned} \tag{4.34}$$

Inserting (4.34) into (4.32) yields the derivative of $\tilde{Q}$ with respect to the angular modes $\bar{\theta}_i$:

$$\frac{\partial \tilde{Q}}{\partial \bar{\theta}_i^{new}} = \sum_{\nu=1}^{n} P^{old}(i|r_\nu) \Bigg\{ \kappa_i^{new}\, \xi_\nu \sin(\psi_\nu - \bar{\theta}_i^{new}) - \kappa_i^{new} \sqrt{1-\xi_\nu^2}\, \cos(\psi_\nu - \bar{\theta}_i^{new}) \tanh\left(\kappa_i^{new} \sqrt{1-\xi_\nu^2}\, \sin(\psi_\nu - \bar{\theta}_i^{new})\right) \Bigg\} \stackrel{!}{=} 0\,.$$

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}:

$$0 = \sum_{\nu=1}^{n} P^{old}(i|r_\nu) \Bigg\{ \xi_\nu \sin(\psi_\nu - \bar{\theta}_i^{new}) - \sqrt{1-\xi_\nu^2}\, \cos(\psi_\nu - \bar{\theta}_i^{new}) \tanh\left(\kappa_i^{new} \sqrt{1-\xi_\nu^2}\, \sin(\psi_\nu - \bar{\theta}_i^{new})\right) \Bigg\}\,. \quad (4.35)$$

Using $\frac{d}{dx} I_0(x) = I_1(x)$ (Abramowitz and Stegun, 1970), the derivative of (4.34) with respect to κ_i^{new} is:

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

Hence, the second update equation is:

$$\frac{\partial \tilde{Q}}{\partial \kappa_i^{new}} = \sum_{\nu=1}^{n} P^{old}(i|r_\nu) \Bigg\{ \frac{I_1(\kappa_i^{new})}{I_0(\kappa_i^{new})} - \xi_\nu\, \cos(\psi_\nu - \bar{\theta}_i^{new}) - \sqrt{1-\xi_\nu^2}\, \sin(\psi_\nu - \bar{\theta}_i^{new}) \tanh\left(\kappa_i^{new} \sqrt{1-\xi_\nu^2}\, \sin(\psi_\nu - \bar{\theta}_i^{new})\right) \Bigg\} = 0\,. \quad (4.36)$$

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, (4.35) and (4.36) 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 maximisation step, e.g., using a Newton-Raphson algorithm.

In the following section it will be shown how the computational cost of each maximisation step can be significantly reduced by introducing closed form approximations of the update equations.

4.2.6 An approximation for closed form update equations

The transcendental nature of the update equations (4.35) and (4.36) 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,

$$\frac{I_1(\kappa_i^{new})}{I_0(\kappa_i^{new})},$$

can be replaced by its asymptotic approximation (Mardia, 1972),

$$\frac{I_1(\kappa)}{I_0(\kappa)} \approx 1 - \frac{1}{2\kappa}. \tag{4.37}$$

Figure (4.5) shows a comparison of the exact ratio vs. the asymptotic approximation. In fact, the approximation is rather "benign" (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 ($\kappa \in [5, 40]$).

The terms with the hyperbolic tangent in equation (4.35) can be simplified by replacing the new parameter values by their values from the previous update step, i.e.,

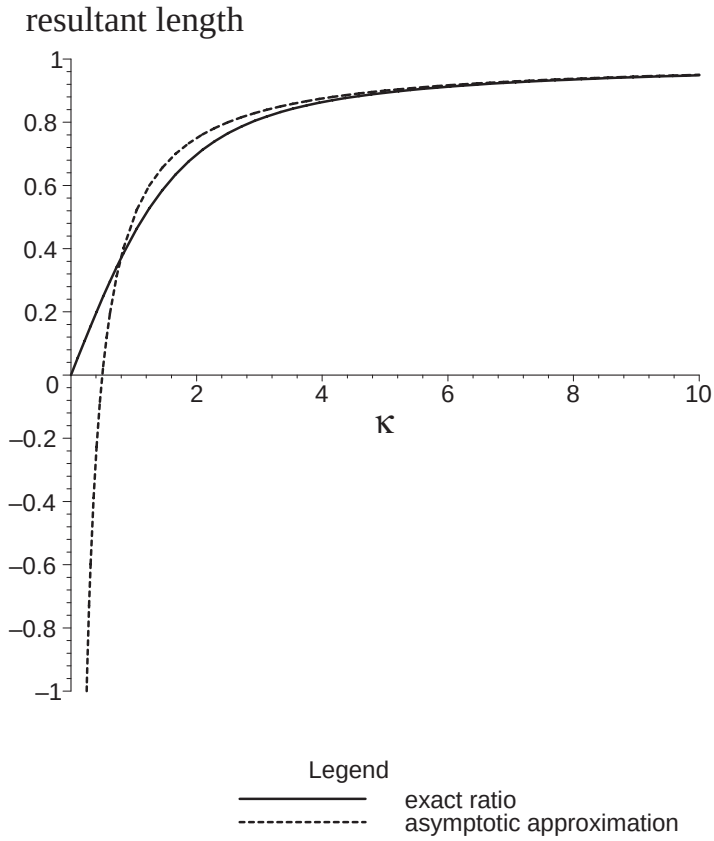


Figure 4.5: Graph of the ratio $I_1(\kappa)/I_0(\kappa)$ (resultant length) and its asymptotic approximation according to eqn.(4.37). For $\kappa > 4.5$ the accuracy is better than 1%. One must, however, ensure that $\kappa > 1$. For all practical purposes, this is no serious limitation.

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

One justification for such an approximation might be the fact that the hyperbolic tangent is a "squashing function" that saturates for arguments greater than one and suppresses small arguments. Thus the term D_ν is basically robust with respect to the exact value of κ_i^{old} and $\bar{\theta}_i^{old}$, as long as the whole argument of the hyperbolic tangent is greater than one. In this case, $D_\nu \approx 1$.

If, on the other hand, the argument is small, then $D_\nu \ll 1$, and the contribution of the entire second term in equation (4.35) is small. Again, the error introduced by "old"

parameter values in D_ν is negligible. An error results only in the few cases when the argument is in the quasi-linear range of the hyperbolic tangent. The approximation is appropriate if the erroneous terms in the summation over ν are in the minority, which depends on the number of data points, n, i.e., the size of the filter population.

Using the abbreviation D_ν for these now purely data-dependent terms, the update equation for $\bar{\theta}_i^{new}$ becomes:

$$0 = \sum_{\nu=1}^{n} P^{old}(i|r_\nu)\left\{ \xi_\nu \sin(\psi_\nu - \bar{\theta}_i^{new}) - \sqrt{1-\xi_\nu^2}\, \cos(\psi_\nu - \bar{\theta}_i^{new})\, D_\nu \right\}.$$

Expanding the trigonometric terms allows solving for $\bar{\theta}_i^{new}$:

$$\begin{aligned} 0 &= \sum_{\nu=1}^{n} P^{old}(i|r_\nu)\left\{ \xi_\nu \Big[\sin\psi_\nu\, \cos(\bar{\theta}_i^{new}) - \cos\psi_\nu\, \sin(\bar{\theta}_i^{new}) \Big] \right. \\ &\qquad \left. - D_\nu\, \sqrt{1-\xi_\nu^2}\, \Big[\cos\psi_\nu\, \cos(\bar{\theta}_i^{new}) + \sin\psi_\nu\, \sin(\bar{\theta}_i^{new}) \Big] \right\} \\ &= \cos(\bar{\theta}_i^{new}) \sum_{\nu=1}^{n} P^{old}(i|r_\nu)\left\{ \xi_\nu \sin\psi_\nu - D_\nu\, \sqrt{1-\xi_\nu^2}\, \cos\psi_\nu \right\} \\ &\quad - \sin(\bar{\theta}_i^{new}) \sum_{\nu=1}^{n} P^{old}(i|r_\nu)\left\{ \xi_\nu \cos\psi_\nu + D_\nu\, \sqrt{1-\xi_\nu^2}\, \sin\psi_\nu \right\} \end{aligned}$$

$$\Longrightarrow \tan(\bar{\theta}_i^{new}) = \frac{\sum_{\nu=1}^{n} P^{old}(i|r_\nu)\left\{ \xi_\nu \sin\psi_\nu - D_\nu\, \sqrt{1-\xi_\nu^2}\, \cos\psi_\nu \right\}}{\sum_{\nu=1}^{n} P^{old}(i|r_\nu)\left\{ \xi_\nu \cos\psi_\nu + D_\nu\, \sqrt{1-\xi_\nu^2}\, \sin\psi_\nu \right\}}. \quad (4.38)$$

Now, there are only data and "old" parameter values on the right hand side.

The simplification of the update equation for the κ_i^{new} is achieved in a similar manner, but 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 (4.37) equation (4.36) becomes:

$$\begin{aligned}
0 &= \sum_{\nu=1}^{n} P^{old}(i|r_\nu)\Bigg\{\left(1-\frac{1}{2\kappa_i^{new}}\right) - \xi_\nu\ \cos(\psi_\nu - \bar{\theta}_i^{new}) \\
&\qquad\qquad - \sqrt{1-\xi_\nu^2}\ \sin(\psi_\nu - \bar{\theta}_i^{new})\,\tilde{D}_\nu\Bigg\} \\
\frac{1}{2\kappa_i^{new}} &= \sum_{\nu=1}^{n} P^{old}(i|r_\nu)\Bigg\{1 - \xi_\nu\ \cos(\psi_\nu - \bar{\theta}_i^{new}) - \sqrt{1-\xi_\nu^2}\ \sin(\psi_\nu - \bar{\theta}_i^{new})\,\tilde{D}_\nu\Bigg\} \\
\kappa_i^{new} &= \frac{1}{2}\,\frac{\sum_{\nu=1}^{n} P^{old}(i|r_\nu)}{\sum_{\nu=1}^{n} P^{old}(i|r_\nu)\Big\{1 - \xi_\nu\ \cos(\psi_\nu - \bar{\theta}_i^{new}) - \sqrt{1-\xi_\nu^2}\ \sin(\psi_\nu - \bar{\theta}_i^{new})\,\tilde{D}_\nu\Big\}}
\end{aligned} \tag{4.39}$$

The approximation sets a lower bound for the value of κ_i, meaning that the angular width of the mixture components must not be too broad. A lower bound for κ_i 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 certainties of the orientation measurement are sufficiently high.

Table 4.1 shows parameter sets obtained through numerical solutions of the full set of transcendental update equations, compared with the results from the approximation. For reasonably high certainty, the approximation is excellent.

a) straight edge, $\kappa_0 = 28.14$	$P(1)$	$P(2)$	$\bar{\theta}_1$	$\bar{\theta}_2$	κ_1	κ_2
full transcendental	1.0	-	5.53°	-	35.91	-
closed form approximation	1.0	-	5.53°	-	35.60	-
b) corner, $\kappa_0 = 28.14$						
full transcendental	0.54	0.46	11.99°	−68.05°	17.69	20.37
closed form approximation	0.54	0.46	11.89°	−67.96°	17.16	19.86

Table 4.1: Comparison of parameters calculated using both the full transcendental update equations and their closed form approximations. The differences are so small that the approximation seems well justified. As expected, the approximation is slightly better for a higher κ.

4.3 Numerical Issues with the EM-algorithm

Even though the EM-algorithm is by no means limited to Gaussian or other standard probability densities, its application to a pdf as unusual as (4.26) requires some caution. The response pdf, $p(r;\psi)$, has two problematic ranges of values where numerical difficulties are expected to occur. One is, obviously, the neighbourhood of the singularities at $r = \pm 1$.

However, the responses never actually reach the values ± 1, unless $p(\theta)$ is a delta function. In practice, $|r| < 1$, due to the effect of the convolution (eqn. 4.7). The likelihood rapidly decreases to "moderate" values in the neighbourhood of the singularity. In fact, in none of the examples studied for this thesis, artificial or natural images, did the responses ever approach the critical neighbourhood of the poles.

A much more severe, but less obviously dangerous, situation arises for responses around value zero, which will be illustrated using the filter responses obtained on a perfect straight edge ($\theta = 0°$) in an artificial image. Figure 4.6 (a) shows the response data from this experiment plotted as points $(r, \log\mathcal{L}[r;\psi])$. The plot reveals one critical

data point (0, 25.14) which has a response value zero (within the numerical precision) and a log-likelihood much higher than the remaining data. This data point, therefore, contributes a substantial part of the total data log-likelihood (the sum of all the individual log-likelihoods). Any potential (numerical) error in its log-likelihood value will introduce a significant error in the total data likelihood and the parameter estimation based on it.

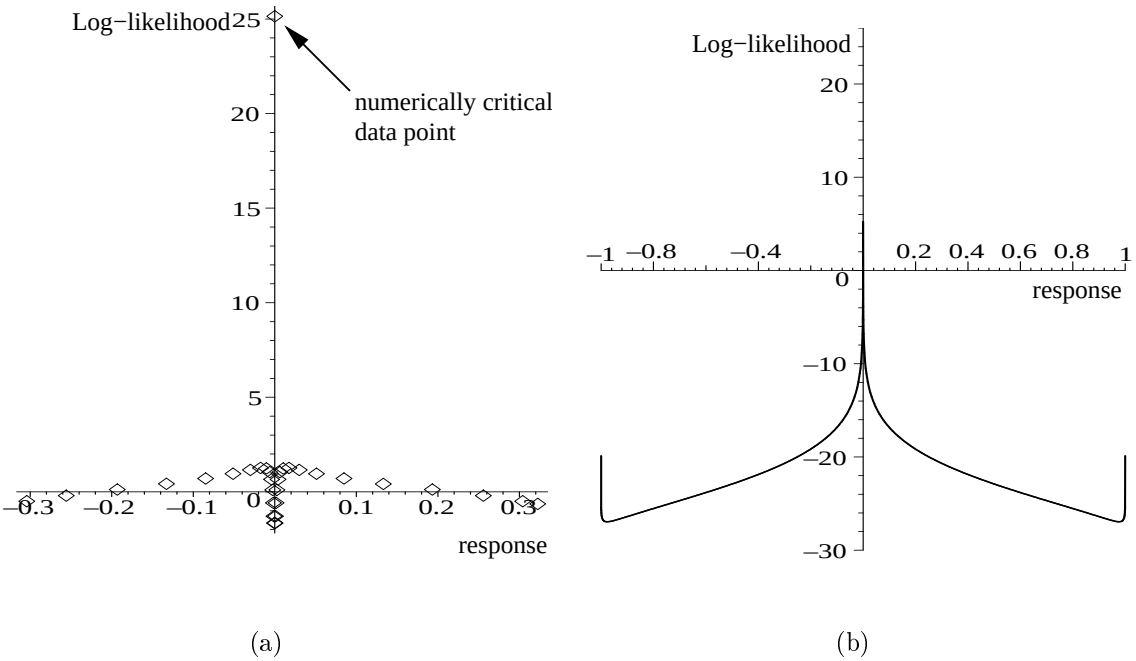
The numerical instability of the log-likelihood function at near zero response becomes more obvious in Figure 4.6 (b), which shows the response log-likelihood for the particular filter (orientation $\psi = 90°$) corresponding to the critical data point in Figure 4.6 (a). The log-likelihood function acts like an amplifier with a strong gain, since the derivative $d\log\mathcal{L}[r;\psi]/dr$ takes high values around $|r| = 0$. Such a system is highly sensitive even to smallest perturbations. The responses are of limited precision due to finite numerical resolution and digitisation, i.e., any response r is given only with a precision $r \pm \Delta r$. Even if $\Delta r \ll 1$, the resultant change in the log-likelihood,

$$\Delta(\log \mathcal{L}) = \frac{d\log\mathcal{L}[r;\psi]}{dr}\Delta r\,,$$

can be extreme. Thus the critical data point is not an "out-lier" in the sense of a false measurement, but rather a measurement with a minute error amplified by the resultant log-likelihood function. It is the finite numerical precision that makes it difficult to process this data point and leads to an error in the total likelihood, and consequently an incorrect parameter estimation.

4.3.1 Avoiding critical data points

The easiest solution is to discard those data points where $|r| \approx 0$, which can be accomplished through a modification of the expectation step (see section.4.2.5) in the EM-algorithm, during which the posterior probabilities are calculated as



(a) (b)

Figure 4.6: **(a)** The response data from an artificial step edge (orientation $\theta = 0°$) plotted as points (r, $\log\mathcal{L}[r;\psi]$). There is one critical data point (0, 25.14) which has a response value zero (within the numerical precision) and a log-likelihood much higher than the remaining data. Therefore, this data point dominates the total data likelihood, and any potential (numerical) error in its likelihood value will have a strong impact on the total data likelihood and the parameter estimation based upon it. **(b)** Graph of the response log-likelihood for the filter (orientation $\psi = 90°$) corresponding to the critical response in (a). Not only is the log-likelihood high around $r = 0$ but also the derivative $d\log\mathcal{L}[r;\psi]/dr$. Thus any small error in r (e.g., due to finite numerical precision) is amplified, and the resultant error of the log-likelihood can be severe. The graph suggests discarding responses within a small interval around zero in order to avoid numerical difficulties.

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

In the update equations (4.30), (4.35) and (4.36), the $P^{old}(i|r_\nu)$ act as weights for the contributions of the n data points. In order to omit the problematic responses with $|r| \approx 0$, the corresponding weight is set to zero and, at the same time, the effective number of data points taken into account is reduced. Let k be the number of data discarded. Then the effective number of data used for the updating, $\tilde{n}$, is simply given by:

$$\tilde{n} = n - k .$$

The decision to discard data is based on a thresholding, and the reduced set of filter responses $\tilde{R}$ is given by:

$$\tilde{R} = \left\{ r_\nu \mid r_\nu > r_{\min},\ \nu \in [1, n] \right\},$$

where $r_{\min}$ is the response threshold (e.g., $r_{\min} = 0.01$) . The weighted summations over all data points,

$$\sum_{\nu=1}^{n} P^{old}(i|r_\nu) \ \ldots , \quad \text{are replaced by} \quad \sum_{r_\nu \in \tilde{R}} P^{old}(i|r_\nu) \ \ldots$$

For instance, the update equation of the priors becomes:

$$P^{new}(i) = \frac{1}{\tilde{n}} \sum_{r_\nu \in \tilde{R}} P^{old}(i|r_\nu) .$$

4.3.2 Initialisation

The initial values for the angular means $\bar{\theta}_i^{(0)}$ are obtained by a peak search in the activity profile of the filter bank, using a 5-neighbourhood; i.e., a particular response r_ν is a peak (local minimum or maximum) if the condition

$$r_\nu = \max(r_{\nu-2}, r_{\nu-1}, r_\nu, r_{\nu+1}, r_{\nu+2}) \tag{4.40}$$

or

$$r_\nu = \min(r_{\nu-2}, r_{\nu-1}, r_\nu, r_{\nu+1}, r_{\nu+2})\,, \tag{4.41}$$

respectively, is satisfied. Thus the number of detected peaks determines the initial number of mixture components, m, which can be modified later during the iterations of the parameter estimation. The filter orientations cover only the range $[0°, \psi = 180°]$. The responses for the higher angles $\psi > 180°$ are given by the antisymmetry relation

$$r(\psi + \pi) = -r(\psi)\,, \tag{4.42}$$

which is a consequence of the sensitivity to contrast sign. In order to check the conditions (4.40) and (4.41) at the "ends" of the filter bank (i.e., responses r_1, r_2, r_n and r_{n-1}), one must compute responses r_{-1}, r_{-2}, r_{n+1} and r_{n+2}, which are not part of the filter bank, by means of (4.42). For a reasonably high number of filters ($n > 16$), this yields an initialisation quite close to the true modes of $p(\theta)$. Let $\psi_{\max,i}$ be the filter orientation corresponding to the i-th peak response. Then,

$$\bar{\theta}_i^{(0)} = \psi_{\max,i}\,.$$

The mixing coefficients $P(i)$ are initialised with the response values at the detected peaks, which have to be normalised by the sum over all response peaks:

$$P(i)^{(0)} = \frac{r_{\max,i}}{\sum_i^m r_{\max,i}}\,.$$

Since there is no easy way of obtaining a first guess for the concentration parameters, they are simply initialised by values similar to those typically obtained from real images, e.g., $\kappa_i^{(0)} = 20$.

Sometimes during the iterations of the EM-algorithm mixture, components "contract" in the sense that their concentration parameters steadily increase beyond any

reasonable value. To avoid such situations, it is useful to impose an upper bound for the concentration parameter and to eliminate a mixture component once that limit is exceeded. The choice of the "concentration threshold" is, of course, heuristic, but, in practice, finding an empirical value that does not erroneously discard necessary components, but which simultaneously limits model complexity, is fairly easy (e.g., $\kappa_{\max} = 100$).

4.4 Measures of Certainty in Probabilistic Population Coding

The probabilistic approach yields not only an estimate for the different edge orientations present in the neighbourhood of the considered point (x, y), but also information about the quality or certainty of these measurements through the concentration parameters of the mixture, the κ_i. A certainty measure, denoted by γ, should be a function of the variance of the angular estimate. Also, it is desirable that the certainty measure be normalised and positive; i.e., $0 \leq \gamma \leq 1$.

However, unlike the case of a Gaussian pdf, where the parameter σ^2 *is* the variance, the corresponding values $1/\kappa_1, \ldots, 1/\kappa_m$ in the von Mises mixture cannot themselves be interpreted as angular variances. In fact, the concepts of mean and variance cannot be uniquely extended to circular statistics; instead they have to be redefined with some care. For an angular variable, the *circular variance*, V_0, is defined as (see e.g., Mardia, 1972):

$$V_0 = 1 - E\{\cos(\theta - \bar{\theta})\}\,, \tag{4.43}$$

where $E\{.\}$ 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

$$\rho = 1 - V_0 = E\{\cos(\theta - \bar{\theta})\}\,, \tag{4.44}$$

which is also referred to as the *resultant length* (Mardia, 1972). 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, one obtains:

$$\rho = E\{\cos(\theta - \bar{\theta})\} = \int_0^{2\pi} p(\theta)\cos(\theta - \bar{\theta})\,d\theta = \frac{1}{2\pi I_0(\kappa)} \int_0^{2\pi} e^{\kappa\cos(\theta-\bar{\theta})}\cos(\theta - \bar{\theta})\,d\theta\,, \tag{4.45}$$

where the integral is essentially the modified Bessel function of the first kind and first order, $I_1(\kappa)$ (Mardia, 1972). Accordingly,

$$\rho = \frac{I_1(\kappa)}{I_0(\kappa)}\,, \tag{4.46}$$

a measure proposed by Zemel and colleagues within the framework of the "directional-unit Boltzmann machine" (Zemel et al., 1995).

4.4.1 Relation to information-theoretic quantities

Information theory provides another important measure of certainty, the Shannon entropy,

$$h = \int_0^{2\pi} p(\theta)\ln p(\theta)\;d\theta\,,$$

which describes the mean information content of a random variable. In statistical estimation theory, the Fisher information,

$$\mathcal{I} = \int_0^{2\pi} \left(\frac{dp(\theta)}{d\theta}\right)^2 \Big/ p(\theta)\;d\theta\,,$$

determines the variance of an *efficient estimator* through the Cramér-Rao (in)equality. However, neither measure is normalised, and the Shannon entropy can even have negative values. The certainty measure ρ introduced in eqn. (4.46) is closely related to the Fisher information, as well as to the Shannon entropy.

For a von Mises distribution, the Fischer information equates to (see appendix A)

$$\mathcal{I}(\kappa) = \frac{1}{2}\frac{I_1(\kappa)}{I_0(\kappa)}\kappa\,.$$

Consequently, ρ is, essentially, the Fischer information scaled by the concentration parameter:

$$\rho = 2\,\frac{\mathcal{I}}{\kappa}\,.$$

The Shannon entropy of the von Mises distribution is given by (see appendix A):

$$h(\kappa) = \ln[2\pi I_0(\kappa)] - \frac{I_1(\kappa)}{I_0(\kappa)}\kappa\,. \tag{4.47}$$

Thus ρ is related to the entropy by:

$$\rho = \frac{\ln[2\pi I_0(\kappa)] - h(\kappa)}{\kappa}$$

4.4.2 A normalised certainty measure based on entropy

Though the resultant length, used by Zemel and co-workers (eqn. 4.46), is a valid certainty measure, it turns out not to be particularly useful for the practical purpose of discriminating local features since it assigns an almost equally high certainty to concentrations $\kappa > 10$. The relevant range of concentration parameters for real images is, however, $5 < \kappa < 40$.

As mentioned in the previous section, certainty may also be defined in terms of the Shannon entropy, but, because the entropy can be negative (for continuous probability

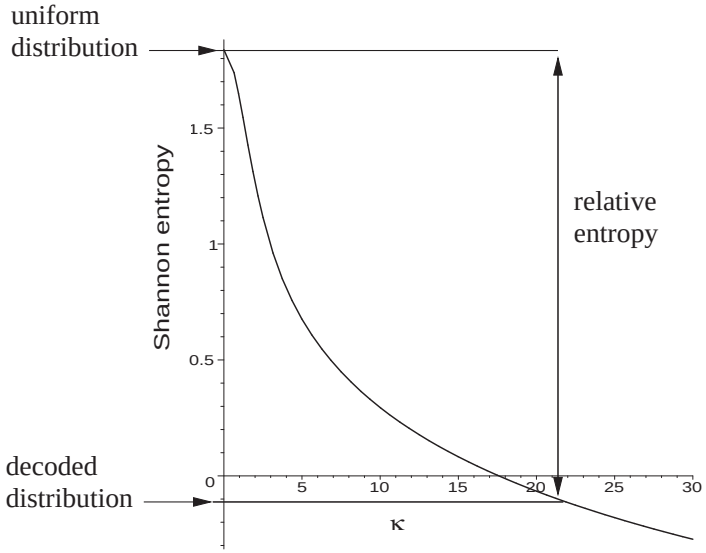


Figure 4.7: Graph of the Shannon entropy of a von Mises distribution as a function of the concentration parameter κ. The maximum $h_{max} = \ln(2\pi)$ is reached for $\kappa = 0$ when the distribution becomes uniform. Note that, unlike for discrete probability distributions, the entropy becomes negative for larger κ-values and has no lower bound. The relative entropy, however, is always positive.

densities), it is unsuitable for practical purposes. In order to derive a *positive* certainty measure from entropy, it is advantageous 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 (see Fig.4.7). Therefore, the Kullback-Leibler divergence,

$$K(p,q) = \int_0^{2\pi} p(\theta) \ln \left[\frac{p(\theta)}{q(\theta)} \right] \, d\theta \,, \tag{4.48}$$

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

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

and, since q is a constant, this yields:

$$K(p_i, \tfrac{1}{2\pi}) = \int_0^{2\pi} p_i(\theta) \left[\ln p_i(\theta) + \ln(2\pi)\right] \, d\theta = -h(p) + \ln(2\pi) = h\left(\tfrac{1}{2\pi}\right) - h(p) \, .$$

With the entropy given by eqn.(4.47), this equates to:

$$\begin{aligned} K(p_i, \tfrac{1}{2\pi}) &= \frac{I_1(\kappa_i)}{I_0(\kappa_i)} \kappa_i - \ln[2\pi I_0(\kappa_i)] \; + \ln[2\pi] \\ &= \frac{I_1(\kappa_i)}{I_0(\kappa_i)} \kappa_i - \ln[I_0(\kappa_i)] \, . \end{aligned} \tag{4.49}$$

Normalisation 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 resultant certainty measure is

$$\gamma = g[K(p_i, \tfrac{1}{2\pi})] = \frac{2}{1 + I_0(\kappa) \, \exp[-\kappa I_1(\kappa)/I_0(\kappa)]} - 1 \, . \tag{4.50}$$

Fig.(4.8) 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 for discriminating measurements by certainty, as it approaches its asymptote of absolute certainty slowly for a very large concentration parameter. The resultant length rapidly approaches the asymptote, already assigning a certainty of $\rho \approx 1$ to κ-values above 10. However, the entropy-based certainty measure allows better judgement of the reliability of orientation measurements in the range relevant for real images, i.e., $5 < \kappa < 40$.

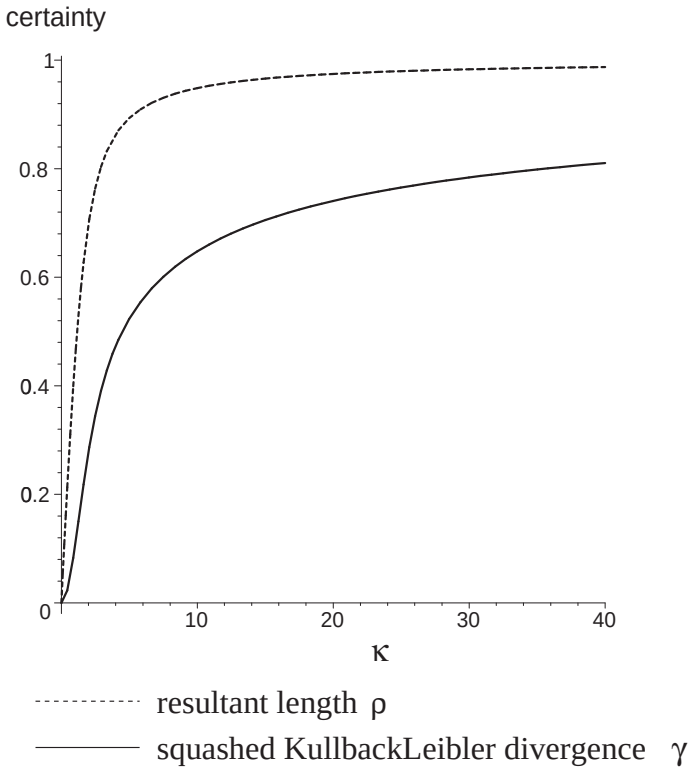


Figure 4.8: Comparison of the resultant length $\rho = I_1(\kappa)/I_0(\kappa)$, proposed as a certainty measure by Zemel and colleagues (Zemel et al., 1995), with the certainty measure γ based on relative entropy (Kullback-Leibler divergence) given by eqn.(4.50). The resultant length (dashed curve) "saturates" too soon whereas γ (solid curve) is more suitable to discriminate between different certainties in the range of $5 < \kappa < 40$ which is important for feature selection in real images.

4.4.3 Average certainty

In order to describe the certainty of a mixture distribution, it makes sense to calculate a weighted average of the certainties of the m mixture components, where the weights are given by the mixing coefficients (priors), $P(i)$, which, by definition, are normalised. This leads to the definition

$$\bar{\gamma} = \sum_{i=1}^{m} P(i)\gamma_i \,. \tag{4.51}$$

The parameter extraction procedure is interpreted as m independent measurements of local orientation, $\bar{\theta}_1, \ldots, \bar{\theta}_m$. Since the average certainty is, by definition, independent of the number of mixture components, it allows comparison between image points, regardless of their orientational structure. A clear corner point, for example, may have a better certainty value than a blurry or noisy edge.

4.5 Experiments on Synthetic and Natural Images

Figure 4.9 shows an image of three objects generating a number of different edge configurations, including straight edges, curved edges, corners and T-junctions. The EM-algorithm described in the previous sections has been applied to a number of key points in this image. For better visualisation, 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 $[180°, 360°)$.

Figure 4.10 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 4.11. 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 4.12 and 4.13) is 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 4.14 and 4.15 have been obtained at 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 principal 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 $\gamma = 0.75$).

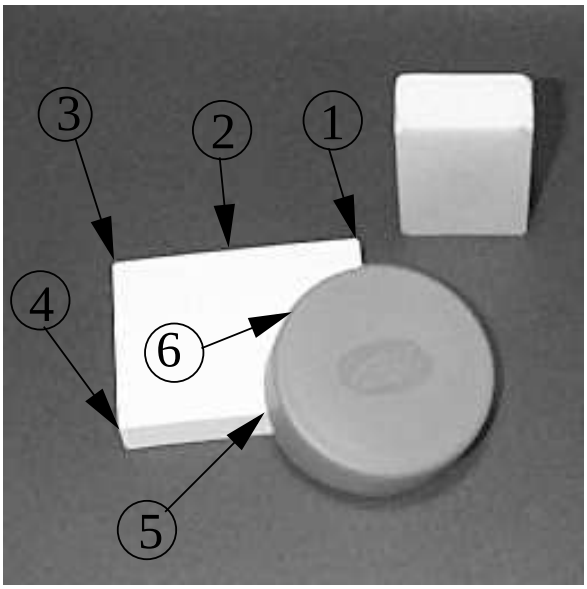


Figure 4.9: A real image with 256×256 Pixels and several points at edges, corners and T-junctions.

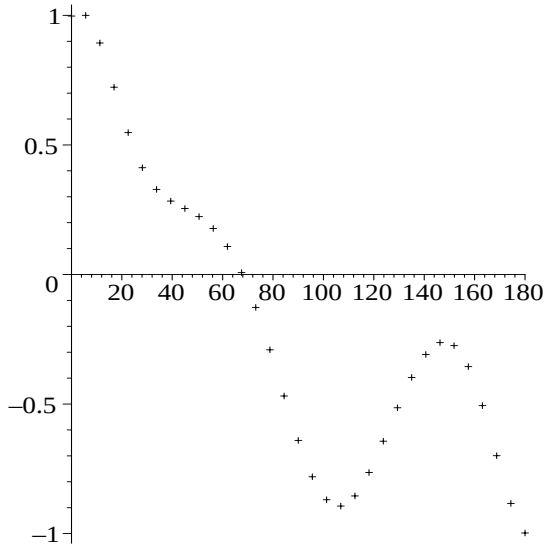


Figure 4.10: Response profile at corner point (1).

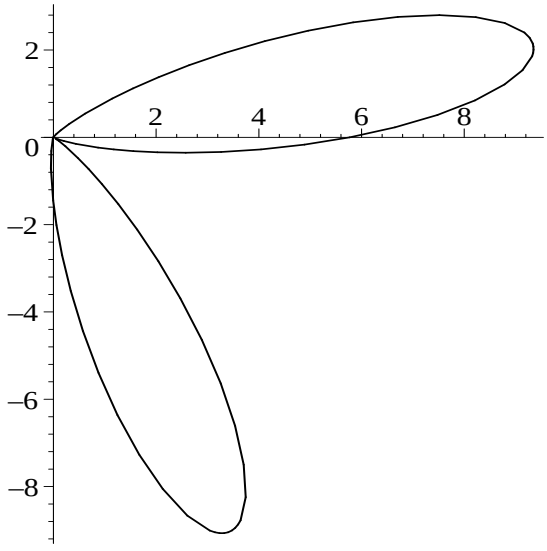


Figure 4.11: Extracted pdf. of local edge orientation in in corner point (1).

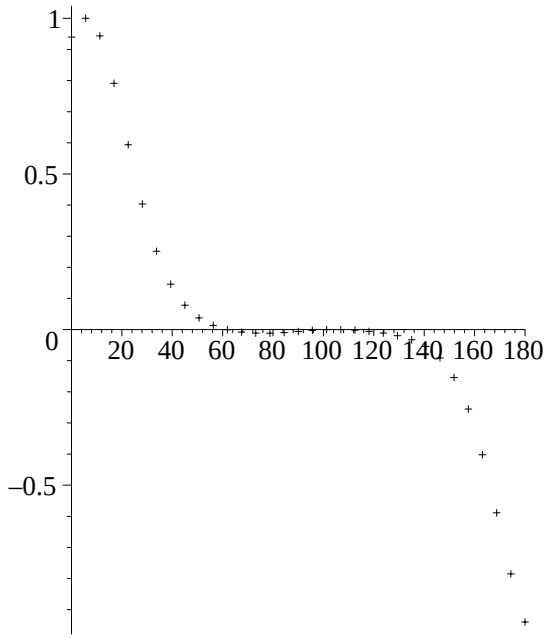


Figure 4.12: Response profile at an edge (point 2).

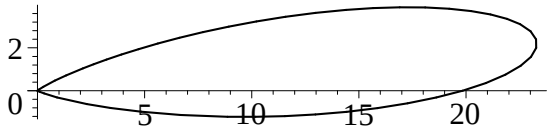


Figure 4.13: Extracted pdf. of local edge orientation at edge (2).

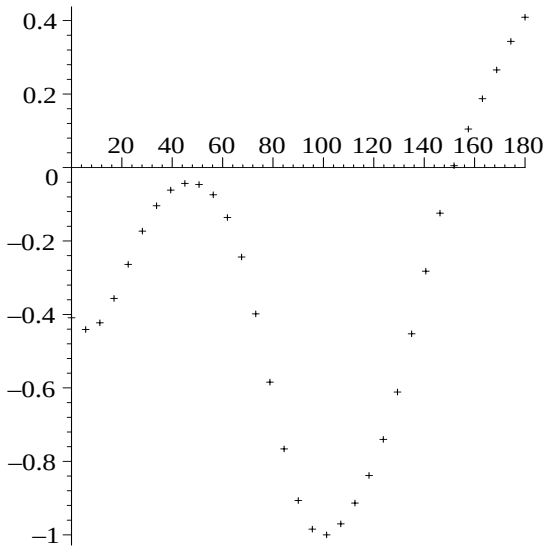


Figure 4.14: Response profile at a T-junction (point 5).

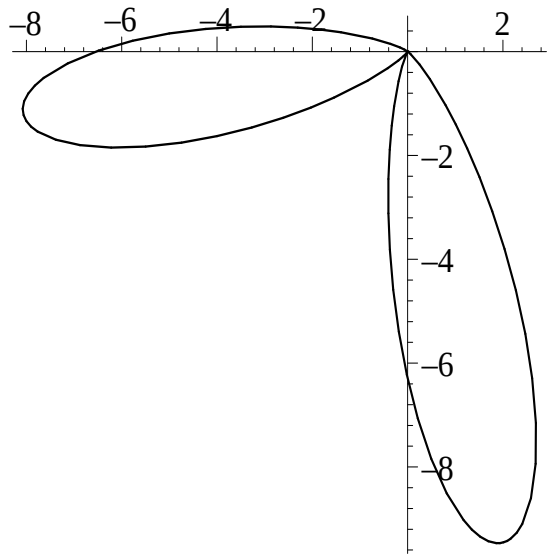


Figure 4.15: Extracted pdf. of local edge orientation in in T-junction (5).

4.5.1 Experiments with varying filter bank size

An important issue in population coding concerns the way in which the estimation error depends on the size of the population. In Figures 4.16 and 4.17, the pdf of local orientation is extracted at points (2) and (5), employing different numbers of filters. The ground truth is represented by dark lines of "manually" determined orientation. Copies of these lines, shifted in parallel, are superimposed with the polar plot of the extracted orientation pdf to enable direct comparison.

The quality of the parameter extraction seems to be affected by the number of filters that are inactive, in the sense that the preferred filter orientation is very different from any local orientation present in the object(s). Nevertheless, inactive filters will pick up noise from the image and contribute this spurious information to the parameter estimation procedure. As illustrated in Section 4.3, the data likelihood can amplify the noise present in small responses. Despite the fact that some precautions are taken in order to discard critical data points, there remains a certain degree of noise introduced by responses that are just above the threshold and that are, therefore, taken into account.

For this reason, at the T-junction the number of filters is not as important as in the case of the straight edge. However, in both cases the variance estimate of the angular distribution is strongly dependent on the number of data points (filters).

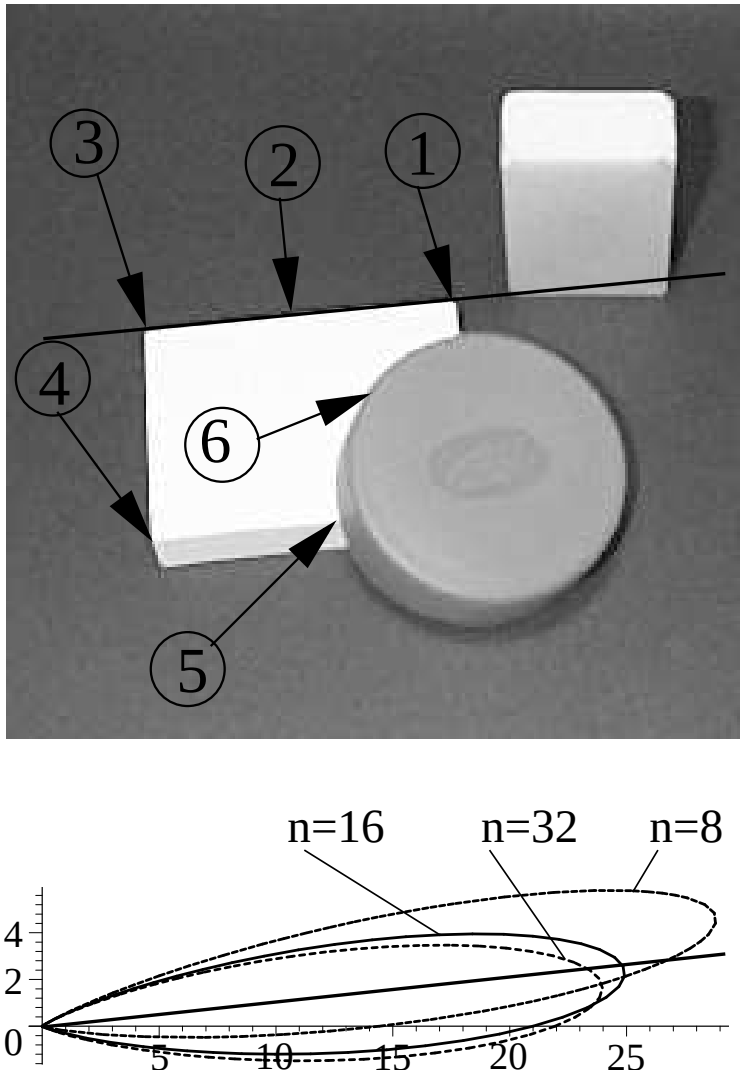


Figure 4.16: The extracted pdf at point (2) for different numbers of filters. Comparison with the ground truth, represented by the dark line, reveals that the orientation estimate is most precise when 16 filters are employed.

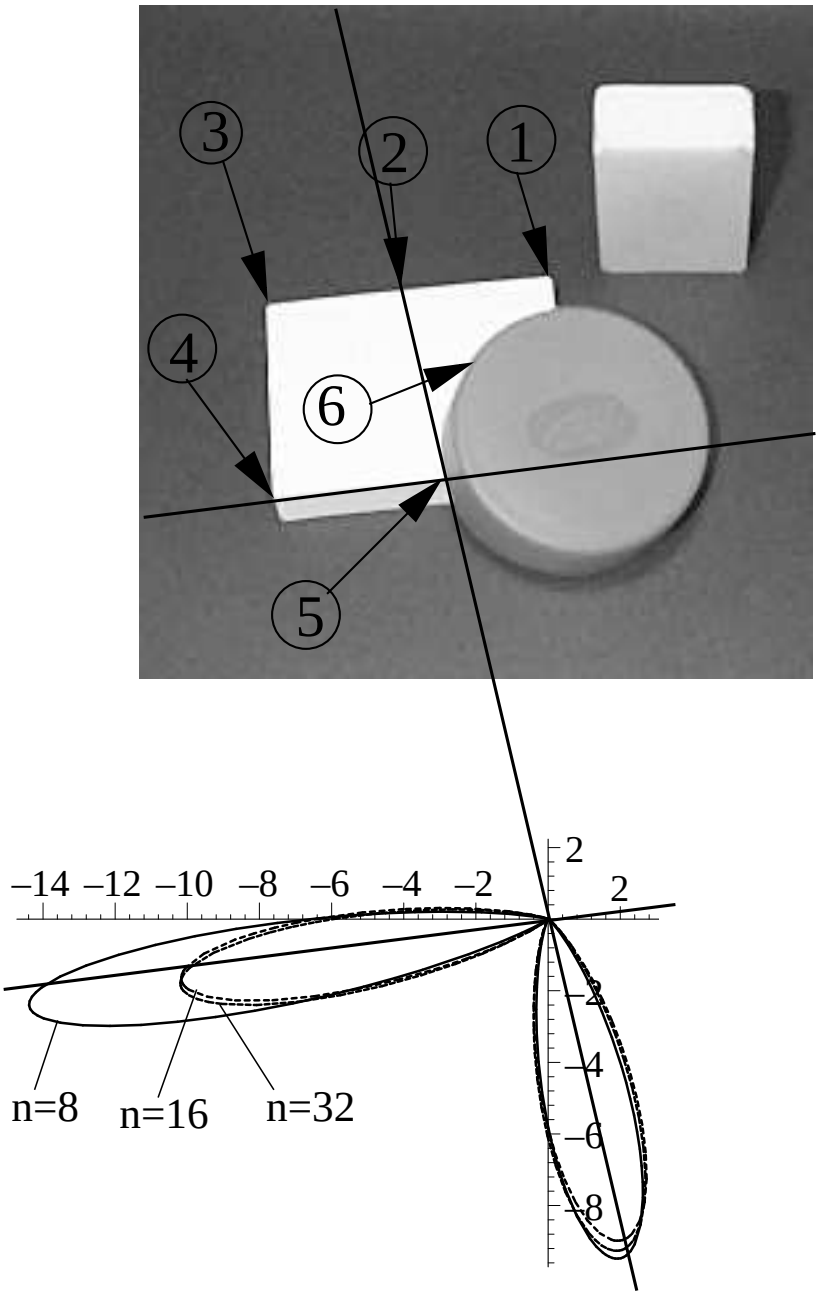
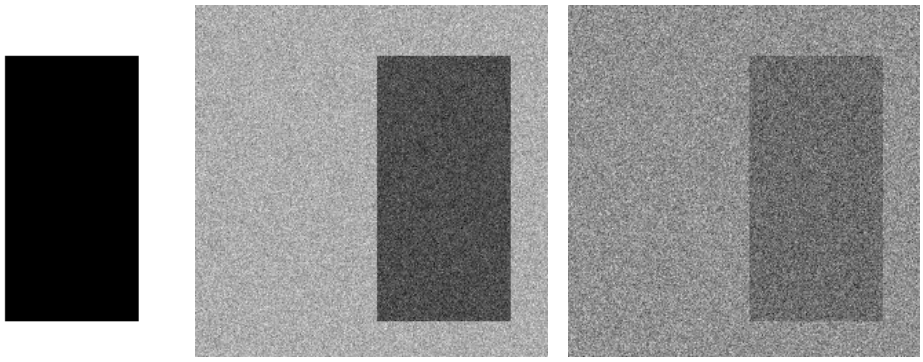


Figure 4.17: The extracted pdf in at point (5) for different numbers of filters. Compared to the ground truth the orientation of the occluded edge is not estimated as precisely as that of the curved edge. Overall, the accuracy of the orientation estimate does not depend very much on the number of filters.

4.5.2 Experiments with noise

In order to study the effect of noise on the certainty measure, it is useful to test the algorithm on an artificial image. A perfect straight edge has been superimposed with Gaussian noise (additive noise) of different variance (Fig. 4.18). The noise level (standard deviation) is given in percent of the maximum contrast. The result for the estimated concentration parameter and the corresponding certainty value is shown in Figure 4.19 for two different filter wavelengths. Both concentration parameter and certainty gradually decrease with increasing noise level. However, for the filter bank with smaller wavelength the performance decreases abruptly at a particular noise level (about 65%, cf. Fig. 4.18 (c)), indicating that the parameter estimation procedure can, at some point, only incorporate the increasing number of response "outliers" by choosing a very large variance (small κ).



(a) (b) (c)

Figure 4.18: (a) An artificial step edge. (b) with added Gaussian noise (20%). (c) with 65% noise, the "breakdown" level of filter bank with $\lambda = 10$ pixels in Fig. 4.19

Effect on multiple orientation estimates

Figure 4.20 shows the influence of Gaussian noise added to the natural image 4.9 on the performance of orientation estimation in the corner point (3).

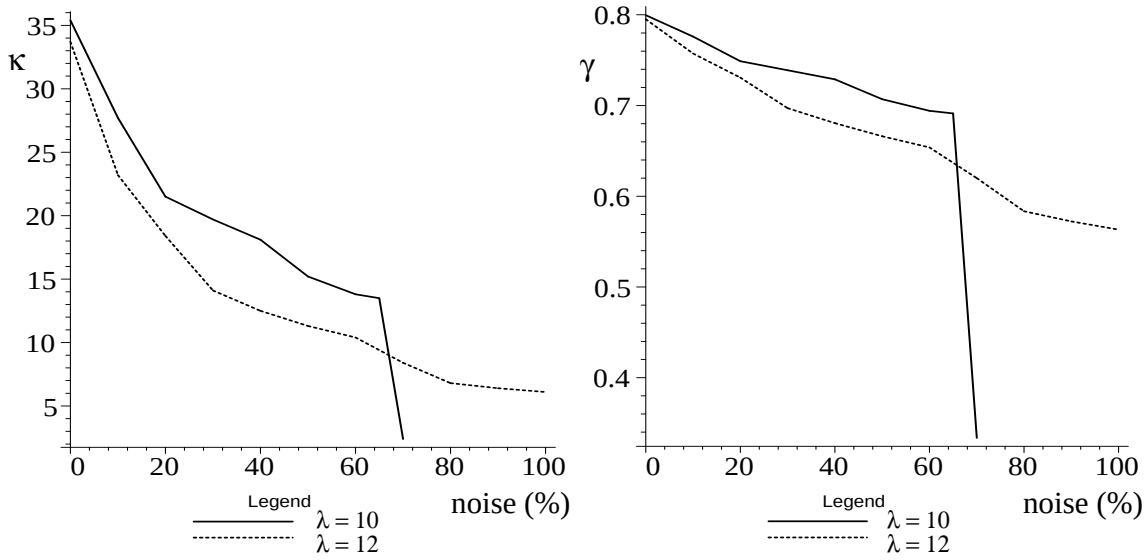


Figure 4.19: An Experiment with Gaussian noise added to an artificial step edge. The plot shows the concentration parameter (**left**) and the certainty (**right**) as a function of the noise level (variance in percent of maximum contrast) for two filter wavelengths. Both concentration parameter and certainty gradually decrease with growing noise level. The filter of smaller wavelength shows an abrupt drop at about 65% noise level. A similar sudden breakdown of performance is expected for the larger wavelength at a higher noise level.

For moderate noise the orientation estimates are reasonably accurate, considering the fact that they are local measurements. However, above 5% noise level the performance deteriorates until the estimates virtually "merge" at about 30% noise. However, it is important to note that the strength of the population coding method is not the noise robustness of the local orientation measurement itself, but rather the ability to accept or discard local measurements based on their certainty. In the following chapter, it will be demonstrated that by incorporating certainty at the local level and at the perceptual grouping level, the noise robustness of contour detection can be improved as well as the robustness against variations in edge contrast.

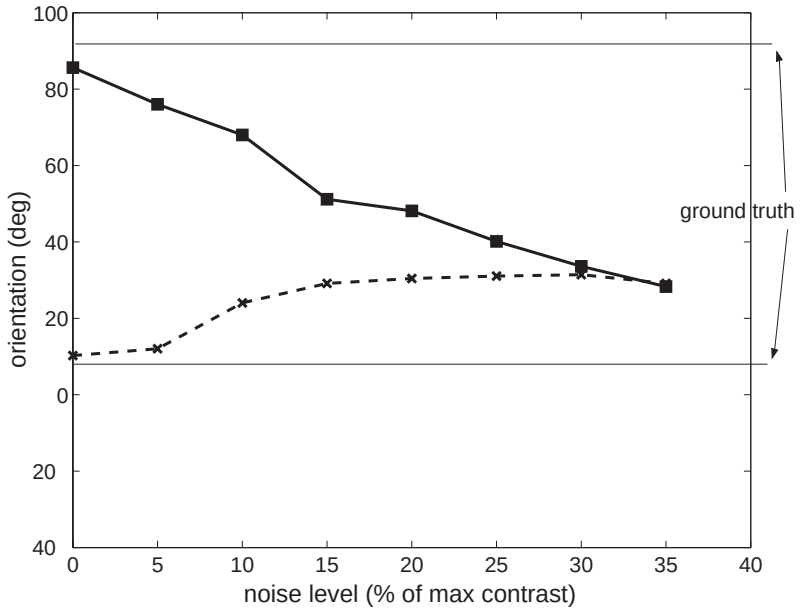


Figure 4.20: An experiment where Gaussian noise was added to a natural image (Fig. 4.9) . The plot shows the mean orientation estimates from the corner point (3) and the corresponding errors. For moderate noise the orientation estimates are reasonably accurate, considering the fact that they are local measurements. However, above 5% noise level the performance deteriorates until the estimates virtually "merge" at about 30% noise.

4.5.3 "Neural" noise

In another experiment, Gaussian noise was added to the filter responses *after* convolution with the image, resembling the situation in a biological neural network, where neural firing rates are stochastic variables. Such an additional noise component is not accounted for by the model, and one cannot expect the system to be very robust against it.

Figure 4.21 shows the influence of additive Gaussian noise on the the mean orientation estimates, obtained by averaging over several hundred trials. The noise standard deviation is given in percent of the maximum response, which in this case is the peak of the stronger component in the response profile in Fig. 4.15:

$$\text{noise level} = \frac{\sigma_{\text{noise}}}{r_{\max}} \,.$$

The noise level can be transformed into the signal to noise ratio (SNR) in decibel (dB) by means of

$$\text{SNR [dB]} = -20 \log(\text{noise level}) \,.$$

The estimate θ_2 corresponds to the weaker component and is, as expected, more affected by the noise than the other component. Obviously, the angular estimators are biased. The mean orientation estimates are roughly stable up to a noise level of 20%, where they start shifting away from their original values. At a noise level of about 40% (SNR $\approx$ 8 dB) of the maximum response, when the noise standard deviation is roughly of the same amplitude as the weaker component (estimate θ_2), the value of $\langle\theta_2\rangle$ saturates, indicating that orientation estimation is no longer functioning.

Similar results are obtained when an orientation estimate is extracted at the straight edge (point 2). Figure 4.22 shows a measured orientation shifting with increasing noise level. The orientation estimate shifts more rapidly with a unimodal response profile

since there are more inactive filters contributing only noise.

In the case of the corner point (1), the response profile consists of two peaks of similar height, though the component with a peak at $\approx 7°$ has a wider "tail" (Fig. 4.10). As in the example of the T-junction, the angular estimates are stable up to about 20% noise level (SNR $\approx$ 14 dB). For higher noise level the estimates, again, shift until the mean estimate for the wider component approaches that of the other component above 40% noise level (SNR $\approx$ 8 dB, Figure 4.23).

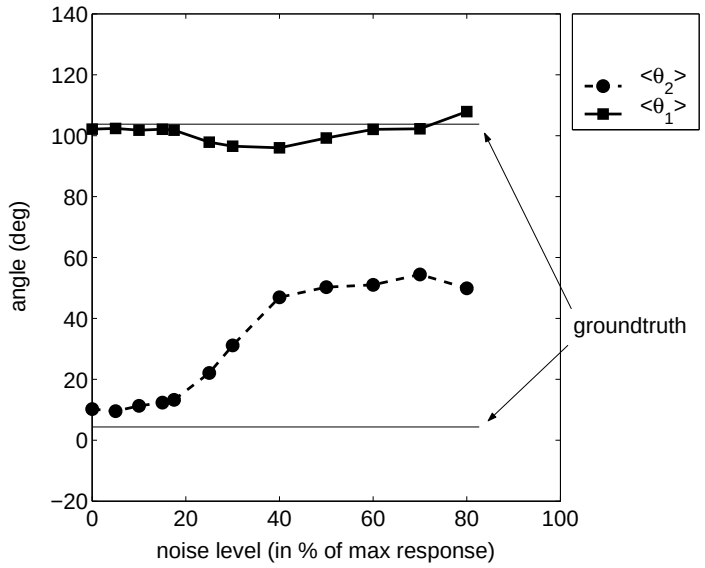


Figure 4.21: An Experiment with Gaussian noise added to the filter responses *after* convolution with the image. The plot shows the mean values of the two orientation estimates,$\langle\theta_1\rangle$ and $\langle\theta_2\rangle$, in the T-junction (point 5) from image 4.9 averaged over several hundred trials, as a function of the noise level (noise standard deviation in percent of the maximum filter response). For comparison, the ground truth is denoted by the dotted straight lines.

4.6 Conclusions and Discussion

In this chapter, a framework has been developed that applies the concept of probabilistic population coding to local edge orientation estimation using a bank of odd-

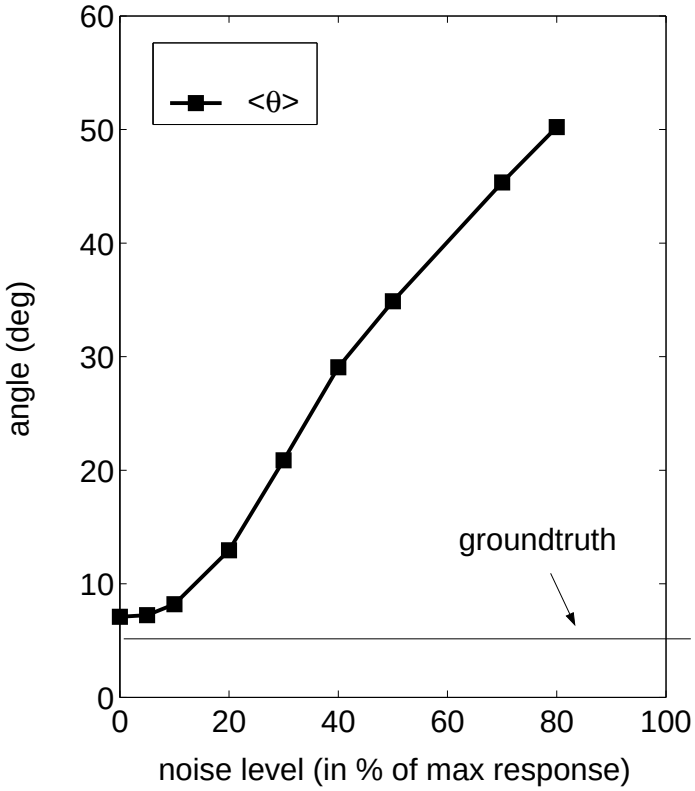


Figure 4.22: An Experiment with Gaussian noise added to the filter responses after convolution with the image. This plot shows the mean value of the orientation estimate, $\langle\theta_1\rangle$, for the straight edge (point 2) from image 4.9 as a function of the noise level (noise standard deviation in percent of the maximum filter response).

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 applied filters, which ensures that responses at two-dimensional edge configurations follow the principle of linear superposition, is essential to the validity of the mixture model.

The fundamental difference of the method presented from the approach of Zemel and colleagues (Zemel et al., 1998) is that the stochastic input variable θ is described

in terms of a parametric model density $p(\theta)$. Moreover, the encoding of the stochastic input variable, θ, is expressed as a transformation of its probability density, $p(\theta)$, into the pdf of the filter responses, $p(r; \psi_\nu)$, which is analytically derived, providing the basis for maximum likelihood estimation of the mixture parameters.

The probabilistic population coding method presented herein constitutes a general feature detector that is capable of representing one- as well as two-dimensional intensity structure using on a set of linear filters. The fundamental difficulties of combined linear detectors for two-dimensional intensity features, identified by Zetzsche and Barth (1990; cf. Section 2.4), do not arise, due to the multiplicative non-linearity of the likelihood function. Thus the likelihood function realizes the non-linear, "and"-like operations postulated by the said authors.

The von Mises model cannot accurately represent certainty in regions with strong curvature, since the "flattened" response profiles obtained in such locations correspond to platykurtic probability densities inherently different from von Mises distributions. As a result, the angular variance is overestimated and consequently the certainty underestimated.

This accords well with the view held by Zucker and co-workers, who consider curvature to be an additional piece of information requiring a separate mechanism of detection, such as "end-stopped" operators (Dobbins et al., 1987). However, incorporating such an elaborate curvature detector would be beyond the scope of this thesis. Curvature will be incorporated in another way in the following chapter, which deals with the grouping of local features by means of interpolating splines.

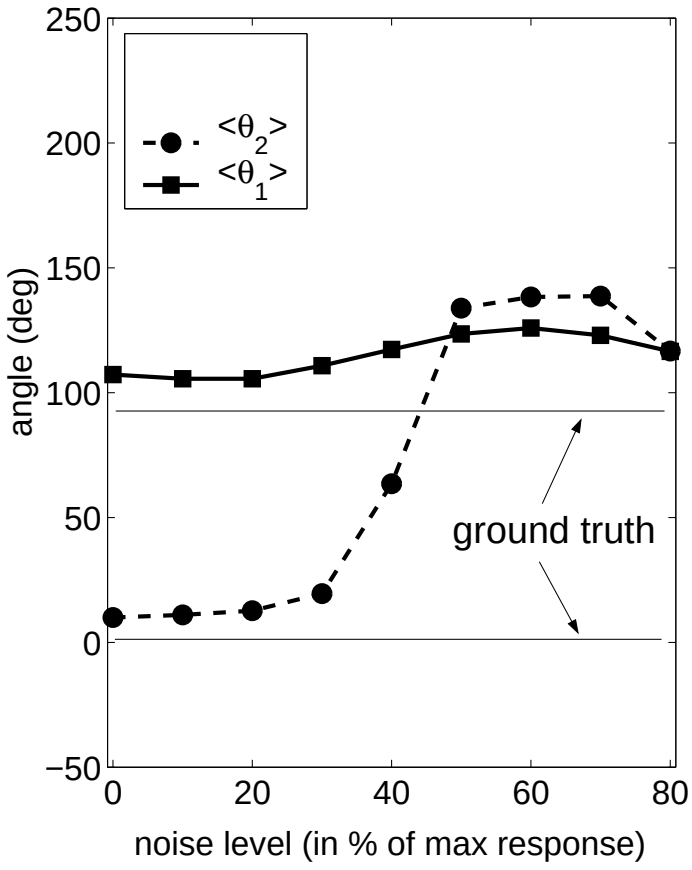


Figure 4.23: An Experiment with Gaussian noise added to the filter responses after convolution with the image. This plot shows the mean value of the orientation estimates $\langle\theta_1\rangle$ and $\langle\theta_2\rangle$ for the corner (point 1) from image 4.9 as a function of the noise level (noise standard deviation in percent of the maximum filter response). The bias with respect to the ground truth (dotted lines) is stronger due to the influence of shadows.

Chapter 5

Contour Segments from Spline Interpolation

Having developed a probabilistic representation of local tangent orientation, it is possible to make a transition from locally identified key points to contour segments by combining the probability densities from pairs of points to determine the parameters of interpolating splines.

Such perceptual organisation is very much in the spirit of Shipley's and Kellman's (1991) psychophysically motivated theory of visual interpolation. Although their criterion of edge *relatability* (see Section 2.7.2 of the literature review) provides a useful test for co-curvilinearity, the geometrical configurations of detected edge segments are in practice often not as precise as required, due to noise. Instead, edge segments would often be erroneously dismissed as "unrelatable" (see Fig. 5.1), particularly on straight contours. Thus orientation estimates will often have to be revised in order to yield "relatable" and more accurate tangent configurations. The revision will be based on mutual consistency, as well as on the quality of agreement between the resultant curve segment and the Gabor transform of the given image. In an abstract sense, the modification of initial orientation estimates can be related to the lateral interactions among

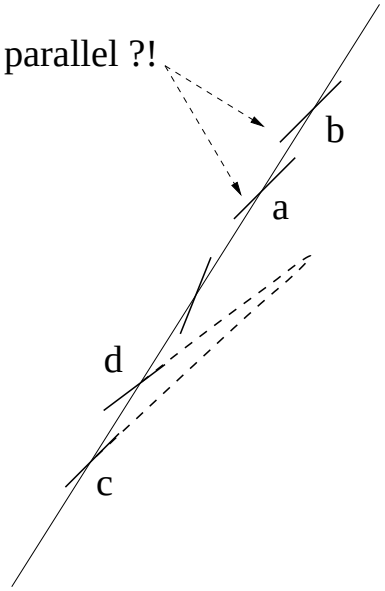


Figure 5.1: Illustration of local orientation "jitter" due to noise. The thin long straight line symbolises a contour, and the thicker short lines are visualisations of extracted local edge orientation. The orientational deviations from the true contour are exaggerated for clarity. Edge segments **a** and **b** appear "unrelatable", since their orientation deviates from the true contour orientation. Even though **c** and **d** can be connected by a curve, they do not fulfill the "relatability" criterion, and the interpolating spline (dashed curve) has no resemblance to the actual contour. Angular smoothing is required to make the edge segments "relatable" and to assign them to one and the same contour. Modification of initial orientation estimates is only justified if the resultant splines provide an improved contour representation.

hypercolumns in the brain, where local orientation sensitivity is altered by the presence of contextual stimuli (cf. Section 2.1.1).

Within the population coding framework, mutual consistency of contour key points can be expressed in terms of a spline likelihood, comprising the joint density of the orientation pdfs and an additional bending constraint. Thus the perceptual grouping step consists of a combination of local population codes.

The spline chosen for interpolation is of quadratic Hermite-type. Its parameters are fully determined by the positions and tangent orientations of the end points, which

is precisely the information provided by the population-coding-based feature detection procedure developed in the previous chapter.

As a consequence of the probabilistic orientation estimation, tangent angles are governed by probability densities, and the uncertainty of tangent orientations is transformed into the system's uncertainty regarding the connecting spline.

An important feature of the perceptual grouping framework presented in this chapter is that the degree of uncertainty in the orientation determines the "inertia" of a local tangent estimate, i.e., how easily an initial orientation measurement (given by a mode in the corresponding pdf) can be modified during consistency optimisation.

In this role of certainty lies a conceptual difference to other grouping schemes. Usually, the coarseness of the initial local orientation measurements is expressed in terms of the likelihood of potential continuations at the grouping level, characterised, for example, by a "support function" in relaxation labelling (Parent and Zucker, 1989; Hancock and Kittler, 1990) or by orientation "votes" propagated through an "extension field" (Guy and Medioni, 1996). The *initial* certainty of the local measurement, however, is not modeled.

In the grouping approach presented in this chapter, local certainties are measured quantities serving as *constraints* during perceptual organisation. Thus by exploiting all the information provided by the extracted probability densities, the algorithm benefits from the richness of the distributed representation of edge orientation.

5.1 Detection of Control Points

Any spline is determined by a set of control points located either on or outside of the curve. Since local orientation is characterised by the extracted probability density at any point on a contour with good accuracy, it is advantageous to place control points *on* the spline. The algorithm then performs spline interpolation based on control point

positions and tangent angles. Points with multiple orientation are assigned more than one tangent angle, and consequently multiple connections to other neighbouring points can be established.

5.1.1 Localisation of edge segments

The procedure to obtain an initial set of *key points* (control points) is an extension of the method of feature selection described in Section 3.6.1.

In order to circumvent the uncertainty relation between precise edge localisation and accurate tangent orientation estimation, two Gabor transforms are performed. Firstly, a filter of higher spatial frequency is chosen to determine response magnitudes for localisation. Secondly, the image is convolved with a low-frequency kernel, providing the basis for population decoding of tangent orientation and the computation of certainty values. While in Chapter 3 (Section 3.6.1) key points are located at lateral local maxima of the product of certainty and the sum of response magnitudes, $\sum_{i=1}^{n} |r_i|$, a different interaction between certainty and response magnitude will be introduced in the following section.

5.1.2 Certainty-controlled normalisation

The fundamental obstacle in trying to segregate genuine contours from noise is the often very poor edge contrast. One means of overcoming the strong variations of local edge contrast is normalising the response value in each pixel by dividing it by the sum of the responses within a small neighbourhood (e.g., 7×7 pixels). However, to avoid an amplification of noise, such a normalisation should be guided by structural constraints that require the presence of an edge-like response profile to activate the normalisation process. The criterion for the decision to activate the normalisation is that the certainty γ exceeds a threshold $\gamma_{\min}$ (e.g., $\gamma_{\min} = 0.5$, cf. Fig. 4.8). If the

certainty remains below the threshold, the filter response is suppressed. Let R be the sum of responses across the filter bank; i.e.,

$$R = \sum_{i=1}^{n} |r_i| \,.$$

Consider a neighbourhood, $\mathcal{N}$, consisting of $n \times n$ pixels. Then, R undergoes the following transformation:

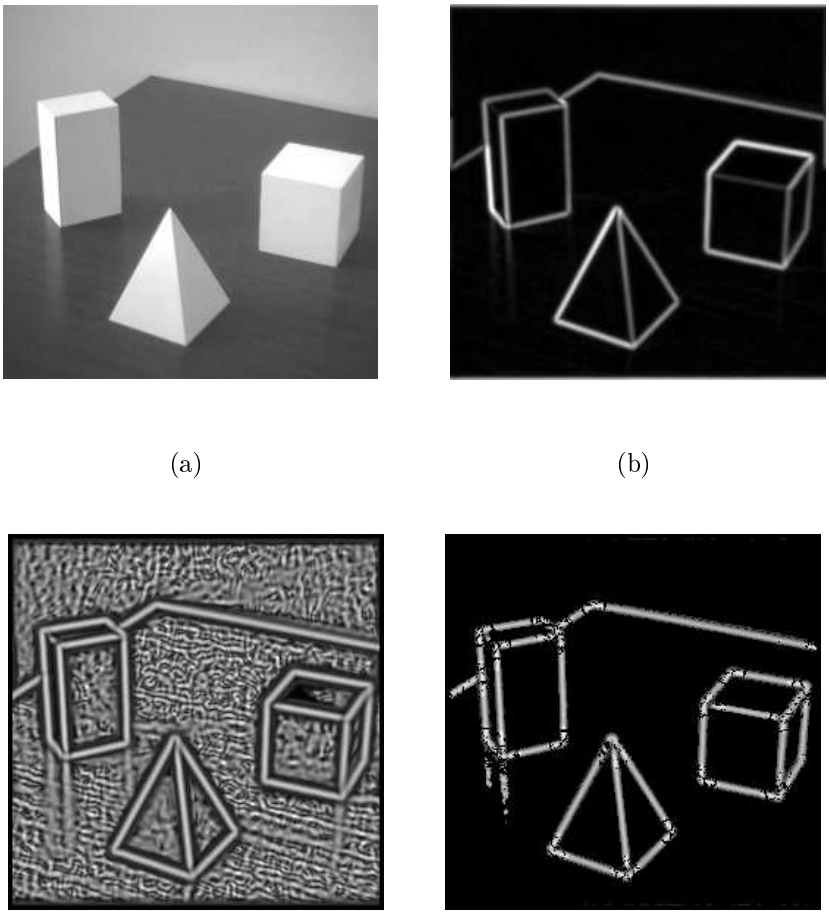
$$R \mapsto \begin{cases} \dfrac{R}{\sum_{j \in \mathcal{N}} R_j} & \text{if} \quad \gamma \geq \gamma_{\min} \\ \\ \dfrac{R}{n^2} & \text{if} \quad \gamma < \gamma_{\min} \end{cases} , \tag{5.1}$$

where the upper row effectively enhances weak contours by levelling out the variations of edge contrast, while the lower, suppresses responses in regions where the response profile does not resemble that of an edge. This facilitative, respectively, suppressive effect of the normalisation increases the signal to noise ratio of the response map.

Normalisation has also been suggested as a biological mechanism *within* cortical hypercolumns by Heeger and colleagues (1996), though in a quite different form and without any reference to certainty. Therefore, the certainty-controlled normalisation is only loosely inspired by their approach.

The choice of the certainty threshold in (5.1) is of general validity, since the typical range of certainties corresponding to object contours in real images is essentially the same for all images, unlike contour contrast, which can be arbitrarily small.

Figure 5.2 illustrates the improved performance of the certainty-controlled normalisation compared to an overall normalisation. Although the desired reduction of response variability is achieved without amplifying noise, the normalisation process also enhances spurious contours in Figure 5.2 (c) that are caused by reflection on the surface of the table. However, reflections of edges are valid structure, since they are regions of genuine "non-accidental" intensity gradient, and a distinction between reflected and "real" edges cannot be made at this level of processing.



(a) (b)

(c) (d)

Figure 5.2: Comparison of a general, "contour-blind" normalisation of the response sum with the more sophisticated, certainty-controlled normalisation. **(a)** Original image. **(b)** The sum of responses ($\lambda = 8$, $\sigma = 0.3$) shows strong variations. **(c)** The response sum, normalised in each pixel (mask size 7×7 pixels). Though responses are levelled out, the signal to noise ratio has become very small. **(d)** The certainty-guided normalisation avoids the extreme amplification of noise that (c) exhibits, by suppressing responses where certainty is low.

5.1.3 Localisation of corner points and junctions

A high-frequency Gabor transform is not suitable for localising points with multiple orientation, since filter responses tend to be less prominent in such locations and stronger in the immediate vicinity. Instead, a measure of multimodality can be derived from the distribution of filter responses with respect to the orientation of the population vector described in Chapter 3.

For this purpose, two new quantities are introduced, which will be referred to as "parallel response integral", $G_{\parallel}$, and "perpendicular response integral", $G_{\perp}$. They are calculated by summing up response *moduli* within the interval of $\pm 45°$ around the population vector and within the interval of $\pm 45°$ around the orientation perpendicular to the population vector. Figure 5.3 illustrates the two angular ranges, using a bimodal response profile.

Let θ_{pop} be the orientation of the population vector (mean orientation), and let $G(\psi)$ be the response profile, i.e., the filter response modulus as a function of the kernel orientation ψ. Then the parallel response integral is defined as:

$$G_{\parallel} = \int_{\theta_{\mathrm{pop}}-\frac{\pi}{4}}^{\theta_{\mathrm{pop}}+\frac{\pi}{4}} G(\psi)\, d\psi\,. \tag{5.2}$$

Analogously, the perpendicular response integral is defined as:

$$G_{\perp} = \int_{\theta_{\mathrm{pop}}+\frac{\pi}{4}}^{\theta_{\mathrm{pop}}+\frac{3}{4}\pi} G(\psi)\, d\psi\,. \tag{5.3}$$

In practice, the continuous definitions (5.2) and (5.3) have to be replaced by discrete approximations:

$$G_{\parallel} = \sum_{i \in \mathcal{I}_{\parallel}} G_i \,, \quad \text{where} \quad \mathcal{I}_{\parallel} = \left\{ i \mid \psi_i \in \left[\theta_{\text{pop}} - \tfrac{\pi}{4}, \theta_{\text{pop}} + \tfrac{\pi}{4} \right] ,\ i = 1, \ldots, n \right\} . \tag{5.4}$$

$$G_{\perp} = \sum_{i \in \mathcal{I}_{\perp}} G_i \,, \quad \text{where} \quad \mathcal{I}_{\perp} = \left\{ i \mid \psi_i \in \left[\theta_{\text{pop}} + \tfrac{\pi}{4}, \theta_{\text{pop}} + \tfrac{3}{4}\pi \right] ,\ i = 1, \ldots, n \right\} . \tag{5.5}$$

One might be inclined to chose the the ratio of the two, $G_{\perp}/G_{\parallel}$, as a measure of multimodality and thus an indicator for the presence of a corner points or a junction. The above ratio is, however, very sensitive to noise, particularly if responses are small. In theory $G_{\perp} \leq G_{\parallel}$, which follows immediately from the definition of $G_{\perp}$ and $G_{\parallel}$: since $G_{\parallel}$ is based on the mean orientation θ_{pop} (orientation of the population vector), at least 50% of the response integral must be concentrated around θ_{pop}. However, the numerator $G_{\parallel}$ can in fact become larger than the denominator, as a result of digitisation. A more robust measure is obtained when the ratio is multiplied by the sum of response magnitudes. Let G be the sum of the magnitudes of all responses in the filter bank,

$$G = \sum_{i=1}^{n} G_i \,.$$

Then, the degree of multimodality, μ, shall be defined as:

$$\mu = \frac{G_{\perp}}{G_{\parallel}} G \,. \tag{5.6}$$

As Figure 5.4 illustrates, μ provides a coarse localisation of corner points and junctions. In particular, the measure is useful to discard certain systematic erroneous corner points created by the intersections of "ripples" in the odd-symmetric Gabor transform (cf. Fig. 5.4 (c)). The rippling effect is quite prominent due to the choice of filter parameters required for good orientation tuning, i.e., relatively large wavelength and size of filter mask.

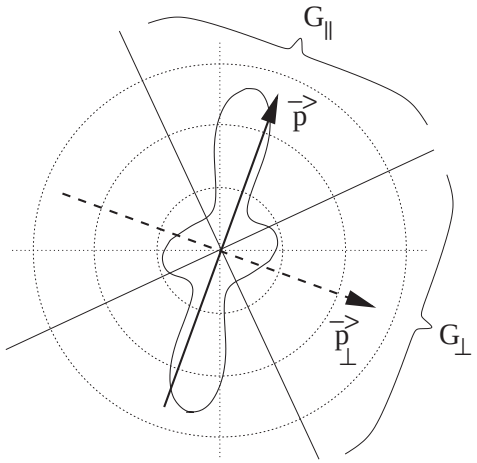
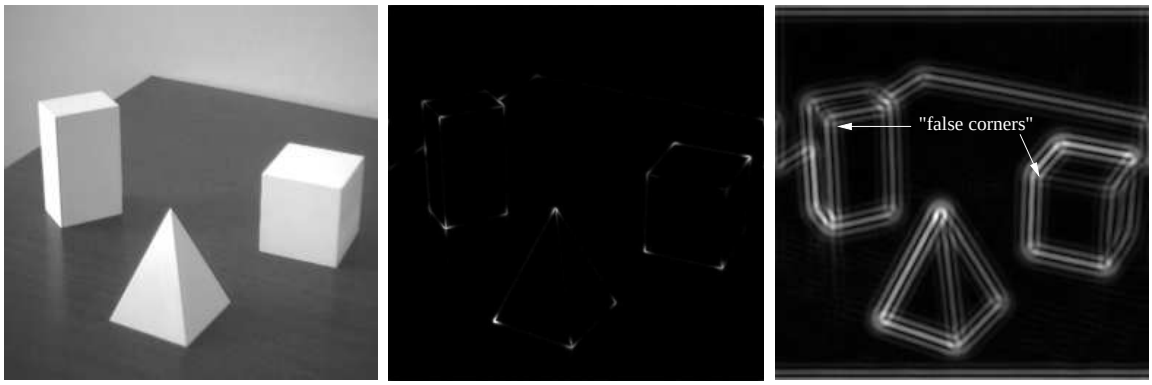


Figure 5.3: Polar plot of a response profile and the corresponding population vector $\vec{p}$ (mean orientation) as well as its perpendicular counterpart $\vec{p}_\perp$. Due to the bimodality of the profile, there is a significant amount of response energy within the $\pm 45°$ range around the orientation of $\vec{p}_\perp$. With a unimodal activity profile almost all the energy would be concentrated around the population vector, unless the tuning curves were extremely wide. A measure of multimodality can be derived from the ratio of the sums of responses around $\vec{p}$ and around $\vec{p}_\perp$.

The multimodality measure only peaks strongly on the contours, since it is computed from phase-insensitive Gabor moduli. An additional focusing effect can be achieved if the response sum G is calculated using the high frequency (localising) Gabor transform.

5.1.4 The uncertainty of key points

The key points are the result of local measurements. Even though they are located on the ridges in the "confidence landscape", defined either by the product of response magnitude and certainty or by certainty-controlled normalisation of response magnitude, there is still a degree of uncertainty, particularly in the tangent orientation, which is crucial, since the orientations measured at the key points will determine the



(a) (b) (c)

Figure 5.4: An image **(a)** and its "multimodality map" **(b)** calculated according to equation (5.6), for a filter wavelength $\lambda = 10$ pixels. The multimodality map, obtained from Gabor moduli, is useful to support the localisation of corners and junction points. **(c)** The map of the odd-symmetrical responses summed over all orientations ($\lambda = 10$ pixels, $\sigma = 0.6\lambda$) contains "ripples", creating "false corners", which can be eliminated by means of the multimodality map.

parameters and consequently the shape of the connecting splines. Therefore, the key points are only preliminary and have to be updated according to smoothness and data consistency constraints. Moreover, spurious key points caused by noise can pass the magnitude-certainty threshold, creating false positives, which can only be identified by their lack of co-curvilinearity with their neighbours at the feature grouping stage.

5.2 Spline Interpolation

An important problem is which type of spline to choose for interpolation between detected key points. B-splines, for example, have become very popular in computer graphics and machine vision, because they are C^2 continuous and naturally represent smoothly curved contours. The C^2 continuity, however, tends to smooth out abrupt

changes of tangent orientation or curvature. An accurate representation of corners is therefore difficult. A B-spline is usually determined only by a set of *external* control points, from which approximate tangent orientations are obtained.

The situation after low-level processing in the population coding approach is different: in addition to the positions of the control points, the corresponding tangent orientations are also known, and points with multiple orientations are identified as such. Therefore, *piecewise Hermite spline interpolation* provides a very straightforward means of connecting such control points.

Hermite splines are a linear combination of a particular type of blending functions, the *Hermite polynomials*. The blending coefficients are the x- and y-coordinates of the control points and the x- and y-components of the corresponding tangent vectors. In the grouping framework presented, the constraint of C^2 and C^1 continuity at the control points will not be imposed. Instead, left and right limits of tangent orientation are introduced, whereby smooth and polygonal curves can be represented equally well (Gavrila, 1996). As a result, the algorithm is capable of representing tangent discontinuities suggested by feature associations, in addition to the locally detected points with multiple orientations.

5.3 Hermite Splines

Once key points and the corresponding probability densities of local edge orientation are available, the parameters of splines interpolating between pairs of key-points can be determined. It is important to note that Hermite splines require the full tangent *vector* at each control point, i.e., the vector of the derivatives of spline coordinates with respect to the spline parameter t. From the orientation pdf, however, only the tangent *angle* is known, not the *magnitude* of the tangent vector. For the simplest form, the *quadratic* Hermite spline, the parameters are fully determined by the end

point positions and the tangent vector in one end point. In this case, however, the unknown magnitude of the tangent vector can be expressed in terms of the tangent orientations in both endpoints. Another motivation for the choice of a quadratic spline is the fact that it has no inflection points, which is in accordance with the concept of "relatability" of curve segments, in particular the "monotonicity constraint" (see Fig. 5.5 (e) and (f)) introduced by Kellman and Shipley (1991).

As a consequence, the inflection points of a contour have to coincide with key points, which is not a demanding requirement, since inflection points are rare and have, by definition, zero curvature and should therefore have a higher certainty than neighbouring points, making them key point candidates *a priori*.

Though the quadratic spline already restricts the possible type of curves, there are ambiguous edge configurations, such as Figure 5.5 (d), where a quadratic spline connection *does* exist, but the required high degree of bending suggests a discontinuity as an alternative. Human perception tends to decide against strongly curved continuations when alternatives with a lesser degree of bending are available (Fig. 5.6).

5.3.1 The quadratic Hermite spline

Let $t \in [0,1]$ be the spline parameter, and let $\mathbf{s}(t)$ denote a position on the spline, i.e., $\mathbf{s}(t) = [x(t), y(t)]^T$. Then the quadratic spline is defined as:

$$\mathbf{s} : [0,1] \longrightarrow \mathbb{R}^2 \,,$$

$$\mathbf{s}(t) = \mathbf{a}t^2 + \mathbf{b}t + \mathbf{c} \,, \qquad \text{with } \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2 \,. \tag{5.7}$$

Here $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ are the vector-valued spline coefficients. Since these quantities are not geometrically meaningful, it is preferable to have the spline defined in terms of endpoint positions and the corresponding tangent angles obtained from population decoding. In

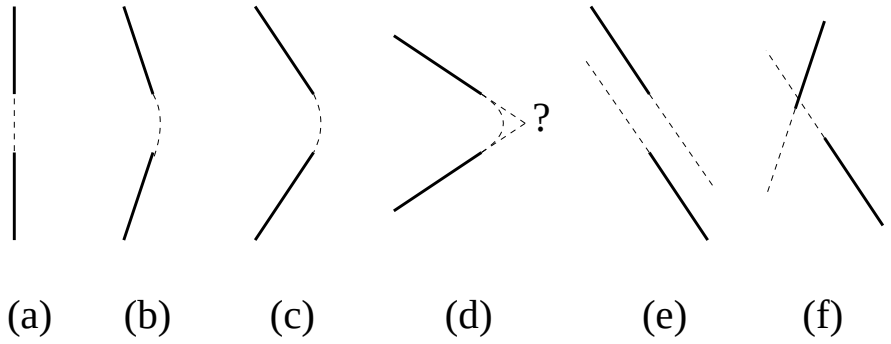


Figure 5.5: A variety of edge configurations. Clearly, (a),(b) and (c) can be connected by smooth curves. Example (d) is ambiguous in the sense that a smooth connecting curve does exist, but the required high degree of bending suggests a discontinuity as an alternative, which tends to be preferred by human perception. Configurations (e) and (f) are intuitively not "relatable", which is reflected in the fact that they would require an interpolating curve with an inflection point. In the terminology of Kellman and Shipley, this is situation is referred to as a violation of the "monotonicity constraint" for interpolating curves.

matrix notation, the spline equation can be written as:

$$\mathbf{s}(t) = [t^2, t, 1] \cdot \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} . \tag{5.8}$$

Accordingly, the first derivative with respect to the parameter t is given by:

$$\dot{\mathbf{s}}(t) = [2t, 1, 0] \cdot \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} . \tag{5.9}$$

Let the position vectors of the endpoints be denoted by $\mathbf{r}_1$ and $\mathbf{r}_2$, and let the tangent vector in endpoint (1) be called $\dot{\mathbf{r}}_1$. These three quantities set the boundary conditions

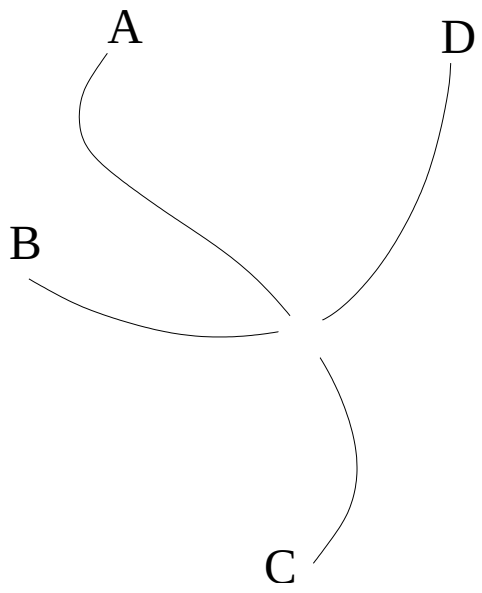


Figure 5.6: Several curved lines with a disrupted (or occluded) centre. Possible associations of line segments are subject to interpretation, but the ambiguity is resolved according to the Gestalt law of "good continuation". Despite the fact that the ends of lines A and B are closer to one another, human perception tends to decide in favour of associating A with C and B with D, avoiding acute angles and strong curvature that would result from any other combination.

of the spline:

$$\mathbf{r}_1 = [x_1, y_1]^T = \mathbf{s}(0)\,, \tag{5.10}$$

$$\mathbf{r}_2 = [x_2, y_2]^T = \mathbf{s}(1)\,, \tag{5.11}$$

$$\dot{\mathbf{r}}_1 = [\dot{x}_1, \dot{y}_1]^T = \dot{\mathbf{s}}(0)\,, \tag{5.12}$$

through which the spline coefficients are completely determined. In matrix notation, this can be expressed as a system of three equations using (5.8) for $t = 0$ and $t = 1$, as well as (5.9) for $t = 0$:

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \dot{\mathbf{r}}_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} . \tag{5.13}$$

This system can be solved for the coefficients:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} = \underbrace{\begin{bmatrix} -1 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}}_{M_H} \cdot \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \dot{\mathbf{r}}_1 \end{bmatrix} . \tag{5.14}$$

By means of M_H, the so-called Hermite matrix, the coefficients can now be expressed in terms of the positions of the end points and the tangent vector. Inserting (5.14) into the spline equation (5.8) yields:

$$\mathbf{s}(t) = [t^2, t, 1] \cdot M_H \cdot \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \dot{\mathbf{r}}_1 \end{bmatrix} . \tag{5.15}$$

Expanding the matrix equation gives the spline in the typical form as a linear combination of $\mathbf{r}_1, \mathbf{r}_2$ and $\dot{\mathbf{r}}_1$, where the corresponding coefficients are given by a set of so-called *blending functions*. In the case of a *cubic* spline, the blending functions are members of the family of *Hermite polynomials*, and the spline is therefore referred to as the *Hermite spline*. Throughout this thesis, the name Hermite spline is adopted for the quadratic spline as well, even though the blending functions are not Hermite polynomials. Hence,

$$\mathbf{s}(t) = (-t^2 + 1)\,\mathbf{r}_1 \; + \; t^2\,\mathbf{r}_2 \; + \; (-t^2 + t)\,\dot{\mathbf{r}}_1 \,. \tag{5.16}$$

Accordingly, the first derivative with respect to t is given by:

$$\dot{\mathbf{s}}(t) = (-2t)\,\mathbf{r}_1 \; + \; (2t)\,\mathbf{r}_2 \; + \; (-2t + 1)\,\dot{\mathbf{r}}_1 \,. \tag{5.17}$$

5.3.2 The magnitude of the tangent vector

Equation (5.16) is a convenient expression of the quadratic spline. However, as mentioned above, the magnitude of $\dot{\mathbf{r}}_1$ is not explicitly given by the probability densities of

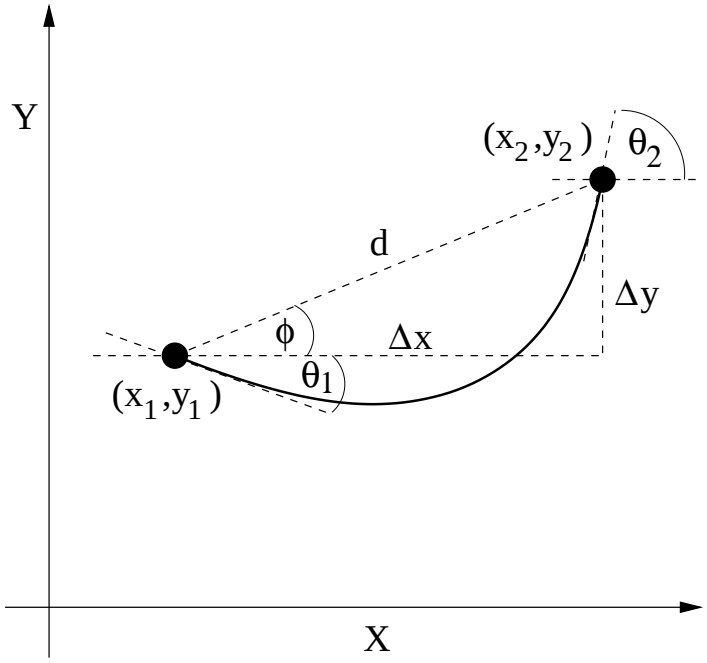


Figure 5.7: A quadratic Hermite spline connecting two points with given tangent orientations θ_1 and θ_2.

orientation obtained from the Gabor filter population code. This lack of information can be compensated by the knowledge of the tangent orientation in the second end point, which is not used in (5.16).

Inserting the additional boundary condition $\dot{\mathbf{r}}_2 = \dot{\mathbf{s}}(1)$ into equation (5.17) yields:

$$\dot{\mathbf{r}}_2 = -2\mathbf{r}_1 + 2\mathbf{r}_2 - \dot{\mathbf{r}}_1 = 2\Delta\mathbf{r} - \dot{\mathbf{r}}_1 \,.$$

This vector equation can be written as two equations in the x and y components:

$$\dot{r}_2 \cos\theta_2 = 2\Delta x - \dot{r}_1 \cos\theta_1 \,. \tag{5.18}$$

$$\dot{r}_2 \sin\theta_2 = 2\Delta y - \dot{r}_1 \sin\theta_1 \,. \tag{5.19}$$

Here θ_1 and θ_2 are the tangent angles in the end points, and $\dot{r}_1$ and $\dot{r}_2$ denote the magnitudes of the tangent vectors. Dividing equation (5.18) by (5.19) eliminates the unknown magnitude $\dot{r}_2$ and leads to an expression of $\dot{r}_1$ as a function of θ_1 and θ_2:

$$\frac{\cos\theta_2}{\sin\theta_2} = \frac{2\Delta x - \dot{r}_1\cos\theta_1}{2\Delta y - \dot{r}_1\sin\theta_1}$$

$$\cos\theta_2(2\Delta y - \dot{r}_1\sin\theta_1) = (2\Delta x - \dot{r}_1\cos\theta_1)\sin\theta_2$$

$$2\Delta y\cos\theta_2 - 2\Delta x\sin\theta_2 = \dot{r}_1(\sin\theta_1\cos\theta_2 - \cos\theta_1\sin\theta_2)$$

$$\Longrightarrow \dot{r}_1 = 2\,\frac{\Delta y\cos\theta_2 - \Delta x\sin\theta_2}{\sin(\theta_1 - \theta_2)}\,. \tag{5.20}$$

Introducing the distance between the end points,

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\,,$$

and

$$\phi = \arctan\left(\frac{\Delta y}{\Delta x}\right),$$

which is the angle of the straight line connecting them, one can write:

$$\Delta x = d\cos\phi \qquad \text{and} \quad \Delta y = d\sin\phi\,,$$

which leads to a more compact expression for $\dot{r}_1$:

$$\dot{r}_1(\theta_1,\,\theta_2) = 2d\,\frac{\sin(\phi - \theta_2)}{\sin(\theta_1 - \theta_2)}\,. \tag{5.21}$$

The above equation immediately elucidates two facts: firstly, $\dot{r}_1$ depends on the distance d between the end points, i.e., the Hermite spline is not scale invariant. Secondly, $\dot{r}_1$ becomes singular for $\theta_1 = \theta_2$. For arbitrary ϕ, this geometric arrangement of tangent vectors and end points would require an inflection point on the spline. Equality of θ_1 and θ_2 is only permissible if simultaneously $\theta_1 = \theta_2 = \phi$. In this case, the spline is a straight line to which $\dot{\mathbf{r}}_1$ and $\dot{\mathbf{r}}_2$ are collinear. However, equation (5.21) still remains

ill-defined. In order to resolve the discontinuity, it is necessary to choose the analytical continuation $\dot{r}_1(\phi, \phi) = 2d$ based on the limit

$$\lim_{\theta_1, \theta_2 \to \phi} \frac{\sin(\phi - \theta_2)}{\sin(\theta_1 - \theta_2)} = 1 \,,$$

which follows from de l'Hopital's rule and results in a meaningful spline (a straight line).

5.4 Optimisation of Spline Parameters

As mentioned, the tangent orientation estimates in the end points, given by the mean orientations of the local mixture components, $\bar{\theta}_i$, are affected by noise, and the relatability criterion should therefore not be applied to such "raw" initial tangent estimates, in order to avoid false negative decisions (see Fig. 5.1). However, the decoded probability densities of tangent orientation represent a range of possible orientations often including a variety of relatable tangent configurations (θ_1, θ_2), even if the initial configuration $(\bar{\theta}_1, \bar{\theta}_2)$ is not relatable.

By virtue of (5.21), each pair $(p(\theta_1), p(\theta_2))$ of two locally extracted orientation densities implicitly represent a density $p(\dot{r}_1 | \, \theta_1)$ in the tangent magnitude $\dot{r}_1$ of the spline, thus describing a "bundle" of possible quadratic splines passing through the fixed end points. There is no need to actually compute $p(\dot{r}_1 | \, \theta_1)$. Instead, the optimisation is performed with respect to the tangent angles θ_1 and θ_2, and the corresponding optimal $\dot{r}_1$ is calculated afterwards.

In order to enforce smoothness of contours, it is necessary to impose a shape constraint on the connecting spline bundle that penalises a high degree of bending. The new tangent angles are then found by means of a maximum likelihood estimation procedure, which results in a trade-off between closeness to initial local measurements and smoothness constraint. The final decision about the relatability of a pair of key points

is made after this optimisation.

5.4.1 The spline likelihood function

In general terms, the total likelihood of a pair of tangent angles (θ_1, θ_2) is given by the product of the joint density of that pair, obtained from population decoding, and a probability density that depends on the degree of bending of the corresponding spline:

$$\mathcal{L}(\theta_1, \theta_2) = p_{\text{pop}}(\theta_1, \theta_2)\, p_{\text{bend}}(\theta_1, \theta_2)\,. \tag{5.22}$$

The quantity that describes the bending of the spline is the magnitude of the second derivative with respect to the parameter t, denoted by $\ddot{s}(t)$. For a quadratic spline $\ddot{s}$ is a constant determined by the boundary conditions. Thus for each end point configuration $(\mathbf{r}_1, \mathbf{r}_2, \theta_1, \theta_2)$ there is a corresponding $\ddot{\mathbf{s}}$. During an optimisation of the angular estimate the end points are kept fixed, leaving the degree of bending a function of the tangent angles alone:

$$\ddot{s} = \ddot{s}(\theta_1, \theta_2)\,.$$

The "bending probability", which acts as a penalty term in the likelihood function (5.22), can be defined as a Gaussian distribution in the scale invariant quantity $\ddot{s}/d$:

$$p_{\text{bend}}(\theta_1, \theta_2) = p(\ddot{s}(\theta_1, \theta_2)) = \frac{1}{\sqrt{2\pi}\sigma_b}\, \exp\left(-\frac{\ddot{s}^{\,2}(\theta_1, \theta_2)}{2d^2\sigma_b^2}\right)\,. \tag{5.23}$$

The variance σ_b^2 determines how strongly bending is penalised and is to be optimised together with θ_1 and θ_2.

In order to find $\ddot{s}(\theta_1, \theta_2)$, equation (5.17) is differentiated once again with respect to t, yielding:

$$\ddot{\mathbf{s}} = -2\mathbf{r}_1 + 2\mathbf{r}_2 - 2\dot{\mathbf{r}}_1 = 2\Delta\mathbf{r} - 2\dot{\mathbf{r}}_1\,.$$

Inserting the definitions (see Fig. 5.7)

$$\Delta\mathbf{r} = d\,(\cos\phi, \sin\phi)^T \qquad \text{and} \qquad \dot{\mathbf{r}}_1 = \dot{r}_1(\cos\theta_1, \sin\theta_1)^T\,,$$

the square of the magnitude can be written as:

$$\ddot{s}^2 = 4(\Delta\mathbf{r} - \dot{\mathbf{r}}_1)^2 = 4d^2 - 8\,d\,\dot{r}_1\,\cos(\phi - \theta_1) + 4\dot{r}_1^2\,.$$

Inserting $\dot{r}_1(\theta_1, \theta_2)$ given by eqn. (5.21) results in:

$$\ddot{s}^2(\theta_1, \theta_2) = 4d^2 - 16d^2\,\frac{\sin(\phi - \theta_2)\cos(\phi - \theta_1)}{\sin(\theta_1 - \theta_2)} + 16d^2\,\frac{\sin^2(\phi - \theta_2)}{\sin^2(\theta_1 - \theta_2)}\,. \tag{5.24}$$

Since the tangent angles θ_1 and θ_2 are independent stochastic variables, their joint density is simply the product of the individual densities. Interpreting an extracted von Mises mixture distribution as a set of superposed measurements each mixture component is treated separately. Consequently, for any pair of points $(\mathbf{r}_1, \mathbf{r}_2)$, the possible spline configurations are described in terms of associations of individual mixture components. The joint probability density of a pair of tangent angles (θ_1, θ_2) using mixture component i from $p(\theta_1)$ and mixture component j from $p(\theta_2)$ is then given by:

$$p_{\mathrm{pop}}(\theta_1, \theta_2 |\, i, j) = p(\theta_1 |\, i)\, p(\theta_2 |\, j) = \frac{1}{4\pi^2 I_0(\kappa_i)\, I_0(\kappa_j)}\, e^{\kappa_i \cos(\theta_1 - \bar{\theta}_i) + \kappa_j \cos(\theta_2 - \bar{\theta}_j)}\,. \tag{5.25}$$

In most cases both points are on an edge, and consequently $i = 1$ and $j = 1$. Only in corner points or junctions several possible associations (i, j) have to be considered. Inserting (5.23) and (5.25) in (5.22) the spline log-likelihood becomes:

$$\begin{aligned} \ln\mathcal{L}(\theta_1, \theta_2;\, i, j) \;\; = \;\; & -\ln[4\pi^2 I_0(\kappa_i)\, I_0(\kappa_j)] + \kappa_i \cos(\theta_1 - \bar{\theta}_i) + \kappa_j \cos(\theta_2 - \bar{\theta}_j) \\ & -\tfrac{1}{2}\ln[2\pi] - \ln\sigma_b - \frac{\ddot{s}^{\,2}(\theta_1, \theta_2)}{2d^2\sigma_b^2}\,. \end{aligned} \tag{5.26}$$

The optimal value of σ_b can be determined by taking the derivative of (5.26) with respect to σ_b and setting it to zero:

$$\frac{\partial}{\partial \sigma_b} \ln \mathcal{L} \;=\; -\frac{1}{\sigma_b} + \frac{\ddot{s}^2}{d^2 \sigma_b^3} \;=\; 0\,,$$

$$\Longrightarrow \sigma_b^2 \;=\; \frac{\ddot{s}^2}{d^2}\,. \tag{5.27}$$

Analogously, setting the derivatives of the log-likelihood with respect to the angles θ_1 and θ_2 to zero results in a set of two transcendental equations for the optimal angles (For clarity, $\ddot{s}^2$ and its derivatives have not been substituted yet):

$$\frac{\partial}{\partial \theta_1} \ln \mathcal{L} \;=\; -\kappa_i \sin(\theta_1 - \bar{\theta}_i) - \frac{1}{2d^2\sigma_b^2} \frac{\partial}{\partial \theta_1} \ddot{s}^2 \;\stackrel{!}{=}\; 0\,. \tag{5.28}$$

$$\frac{\partial}{\partial \theta_2} \ln \mathcal{L} \;=\; -\kappa_j \sin(\theta_2 - \bar{\theta}_j) - \frac{1}{2d^2\sigma_b^2} \frac{\partial}{\partial \theta_2} \ddot{s}^2 \;\stackrel{!}{=}\; 0\,. \tag{5.29}$$

By inserting equation (5.27) the unknown σ_b^2 can be expressed in terms of $\ddot{s}^2$. Before substitution, equations (5.28) and (5.29) are multiplied by σ_b^2. The result is a system of equations only in θ_1 and θ_2 which has to be solved numerically:

$$\kappa_i \sin(\theta_1 - \bar{\theta}_i) \frac{\ddot{s}^2(\theta_1, \theta_2)}{d^2} + \frac{1}{2d^2} \frac{\partial}{\partial \theta_1} \ddot{s}^2(\theta_1, \theta_2) \;=\; 0\,. \tag{5.30}$$

$$\kappa_j \sin(\theta_2 - \bar{\theta}_j) \frac{\ddot{s}^2(\theta_1, \theta_2)}{d^2} + \frac{1}{2d^2} \frac{\partial}{\partial \theta_2} \ddot{s}^2(\theta_1, \theta_2) \;=\; 0\,. \tag{5.31}$$

The bending quantity $\ddot{s}^2(\theta_1, \theta_2)$ is given by equation (5.24), and its derivatives with respect to the angles are:

$$\frac{\partial}{\partial\theta_1}\ddot{s}^2 = \frac{16d^2\sin(\phi-\theta_2)\cos(\phi-\theta_2)\sin(\theta_1-\theta_2) - 32d^2\sin^2(\phi-\theta_2)\cos(\theta_1-\theta_2)}{\sin^3(\theta_1-\theta_2)}\,. \tag{5.32}$$

$$\frac{\partial}{\partial\theta_2}\ddot{s}^2 = \frac{-16d^2\cos(\phi-\theta_1)\sin(\phi-\theta_1)\sin(\theta_1-\theta_2) + 32d^2\sin(\phi-\theta_2)\sin(\phi-\theta_1)}{\sin^3(\theta_1-\theta_2)}\,. \tag{5.33}$$

As an alternative to solving the system (5.30) and (5.31) by means of a Newton-Raphson algorithm, a simple gradient descent can be applied to find a a pair of angles that maximises the log-likelihood function. A straightforward initialisation of the angles is given by the modes $\bar{\theta}_i$ and $\bar{\theta}_j$ of the extracted densities:

$$\theta_1(0) = \bar{\theta}_i \qquad \text{and} \qquad \theta_2(0) = \bar{\theta}_j\,.$$

In most cases these initial values are close to the optimal solution so that convergence of the gradient descent is unproblematic. However, due to the denominators in equations (5.32) and (5.32), numerical problems are expected when the tangent angles θ_1 and θ_2 are very similar. This corresponds to the situation discussed earlier in Section 5.3.2, where the spline takes the form of a straight line.

When numerically large values do not cancel each other out in equations (5.32) and (5.33), the derivatives are computed incorrectly and so is the gradient of the likelihood given by equations (5.28) and (5.29). As a result, the gradient descent procedure tends to move away from the initial point and from the global maximum, and converges to a distant local maximum, which usually is not an improvement over the initial degree of bending. In other words, the optimal straight-line-solution ($\theta_1 = \theta_2 = \phi$) is surrounded and shielded by a "numerically impenetrable barrier", which prevents the algorithm from converging towards it.

A good strategy is therefore to check whether the solution found by the optimisation process has actually led to an improvement and to enforce the straight-line-solution otherwise.

5.4.2 The role of certainty during parameter optimisation

It is important to note that in the log-likelihood function (5.26) the concentration parameters κ_i and κ_j of the pdfs act as weights of angular modifications during the optimisation process. If a concentration parameter is large, any deviation from the initial orientation $\bar{\theta}$ will result in a sharp decrease of the likelihood function unless the overall curvature is substantially reduced simultaneously. In other words, the concentration parameters (and thus the certainties, which are monotonic functions thereof) determine the "inertia" of orientation estimates, i.e., their "flexibility to compromise for the sake of mutual consistency". Herein lies the essential difference to other grouping methods, where measurement of certainty is not an integral part of local feature extraction.

5.5 Organisation of Local Features

All edge key-points obtained through competitive feature selection are stored as nodes in a graph. For each key point, only a limited number of its nearest neighbours are considered for grouping, reflecting the Gestalt law of *proximity*. Having optimised the interpolating splines to the k nearest neighbours, a choice has to be made which of these possible connections to regard as the best representation of the local contour segment(s).

Relatability is only a geometric relation between the tangents in the end points of a contour segment. It does not include any information about the consistency with the image at intermediate points on the interpolating curve. The higher the noise level, the

higher the number of detected false positive key points, and the more likely it becomes that pairs of edge segments are "relatable by chance". Therefore, the decision to associate two key points and to connect them with a spline should not merely depend on relatability of endpoint tangents, but also on the consistency between the image intensity distribution and the spline as a whole.

5.5.1 Spline consistency

The task to evaluate the consistency between the different contour hypotheses, provided by the splines from a key point to its k neighbours, and the data from the Gabor transform is equivalent to the search for most perceptually "salient" curves.

Cross and Hancock (1999) applied a strategy that sums up the response energy along the spline and divides the result by the curve length. Let a point on the spline be denoted by $\mathbf{s}(t)$ and the response value at a particular point, by $R[\mathbf{s}(t)]$. According to their definition, the spline consistency $\mathcal{C}$ is then the integral of response energy over the entire parameter range $t \in [0, 1]$ scaled by the curve length l:

$$\mathcal{C} = \frac{1}{l} \int_0^1 R[\mathbf{s}(t)]\, dt\,,$$

which in practice is approximated by a discrete sum.

A drawback of this approach is that it collects response energy regardless of how the energy in each point is distributed across the filter bank, i.e., without taking into account the directional structure of the detected intensity gradient. Consequently, the method will sum up energy, even if a spline crosses a contour or when it simply passes through a region with strong noise but no genuine contour. Furthermore, the above consistency measure is, by definition, contrast dependent. Thus a faint contour will yield a smaller sum of energy, even if the spline matches the contour very accurately.

Therefore, it is argued that perceptual saliency of an image region should be defined

in a contrast independent manner. According to the philosophy of this thesis, this will be accomplished by taking into account the distribution of filter responses in orientation space, in addition to their distribution in the image plane.

Through its parametric derivative, a spline provides a tangent orientation in every point, which can be compared with the actual orientation measurement obtained from probabilistic population coding. Instead of merely collecting response strength along the spline, it is more efficient to check the *tangent consistency*, i.e., how well the locally measured orientation pdf actually matches the tangent angle of the spline. Hence, the new consistency measure sums up the cosines of the difference between the tangent orientation $\theta(t)$ of the spline, given by

$$\theta(t) = \arctan\left[\frac{s_y(t)}{s_x(t)}\right] ,$$

and the local orientation estimate, $\bar{\theta}(t)$, weighted by the corresponding certainty, $\gamma(s)$:

$$\mathcal{C} = \frac{\int_0^1 \gamma(s) \left| \cos\left[\theta(s) - \bar{\theta}(s)\right] \right| ds}{\int_0^1 \gamma(s)\, ds} \approx \frac{\sum_j^N \gamma_j |\cos(\theta_j - \bar{\theta}_j)|}{\sum_j^N \gamma_j} , \tag{5.34}$$

where the denominator ensures normalised weights and, consequently, independence of curve length.

The latter part of equation (5.34) is essentially a modified tangent log-likelihood of the spline, given the local von Mises-type orientation pdfs extracted on the curve:

$$\ln \mathcal{L} = \ln \prod_j^N e^{\kappa_j \cos(\theta_j - \bar{\theta}_j)} = \sum_j^N \kappa_j |\cos(\theta_j - \bar{\theta}_j)| . \tag{5.35}$$

Apart from the normalisation of weights, the essential difference to (5.34) is that the concentration parameters κ_j in (5.35) are substituted by the certainties γ_j (monotonic functions of the κ_j).

Thus the spline-data consistency measure is an abstraction from absolute contour contrast, as it only relies on relatability, certainty and orientational consistency.

As a final modification, equation (5.34) is multiplied by the ratio of the number of points on the spline where the algorithm was unable to detect an edge-like response profile, M, to the total number of points on the discretised spline, N. Hence the consistency measure becomes:

$$\mathcal{C} = \frac{M}{N} \frac{\sum_{j}^{N} \gamma_j |\cos(\theta_j - \bar{\theta}_j)|}{\sum_{j}^{N} \gamma_j} \tag{5.36}$$

The consistency value $\mathcal{C}$ of a spline connecting two key points serves as the edge label of the connection between the corresponding nodes in the spline graph representation. Connections with poor consistency can then be eliminated by setting a threshold, in order to reduce the connectivity of the graph.

Another simple "pruning" strategy is the elimination of isolated individual splines connecting pairs of spurious key points. Key points that are connected to only one partner are highly likely to be "false positives" due to noise, and should therefore be discarded.

5.5.2 The contour representation algorithm

All the ingredients of the contour detection procedure have been introduced. In summary, the different steps of the algorithm are:

- Gabor filtering
- Extraction of orientation pdfs by Probabilistic population decoding
- Certainty-controlled response normalisation

- Compute multimodality map
- Detect Key-points through feature competition
- Determine the k nearest neighbours to each key point
- Check Kellman-Shipley-relatability with all k neighbours
- Optimise tangent angles to improve mutual feature consistency
- Again: check Kellman-Shipley-relatability with all k neighbours
- Determine spline consistency with image data (before and after angular optimisation)
- Discard those key points that are not relatable to any neighbour
- Apply simple pruning operations (e.g., consistency thresholding)
- Draw most consistent spline(s)

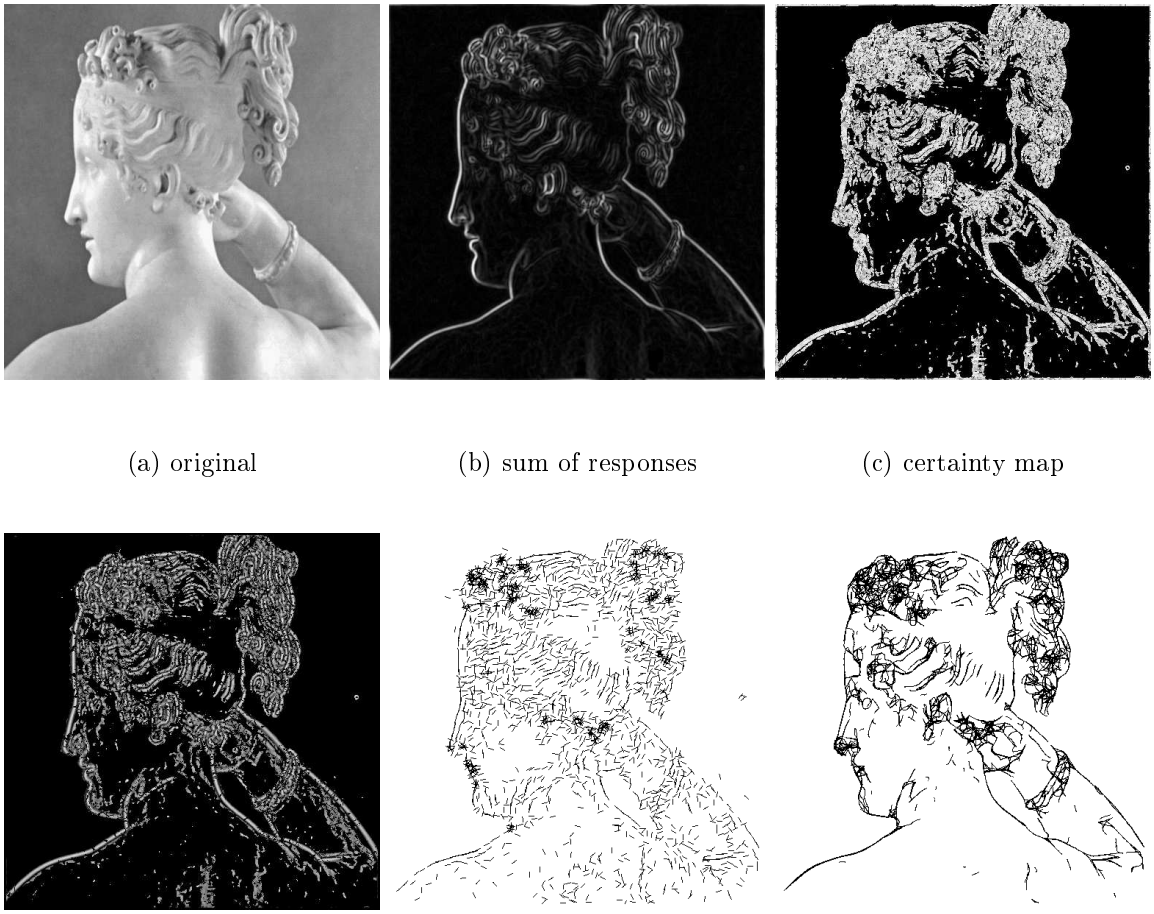
5.6 Experiments

Figure 5.8 shows an image of part of the sculpture of Paolina Borghese by Antonio Canova (1757-1822) and the different steps of contour extraction, from Gabor responses to the tangent elements extracted at key points and the result of spline interpolation between them. Note that the certainty-controlled normalisation effectively combines the characteristic properties of normalised response sum (Fig. 5.8 (b)) and certainty map (Fig. 5.8 (c)). Nearby edges are more clearly separated and weak contours are enhanced. Some parts of the hair region would, however, require a line detector, rather than an edge detector. Thus one cannot expect the system to represent *all* relevant linear details.

The same image has been used by Iverson and Zucker (1995) to demonstrate the performance of their "logical/linear operators". Referring to earlier work by Koendrink and co-workers (1982), the authors point out the perceptual significance of bifurcations and line terminations in regions, such as the folds around the neck, which provide vital information about three-dimensional structure. They also demonstrate that the Canny detector (Canny, 1986), like any other essentially linear edge operator, is not capable of correctly representing bifurcations and tends to smooth out tangent discontinuities in corner points and T-junctions.

Both logical-linear operators and certainty-controlled normalisation impose structural constraints that can deactivate the response of a linear operator. While a logical-linear operator requires that conditions for the existence of an edge or line-like structure are met within its spatial support, certainty controlled normalisation uses response consistency across the filter ensemble as the criterion for activating the operator. It is the rigorous deactivation of responses in regions without edge structure that enables a better distinction of nearby contours than with purely linear filtering techniques.

Since feature extraction with probabilistic population coding explicitly represents points with multiple orientations and orientation discontinuities, the spline interpolation algorithm can accurately capture most of the essential discontinuities and bifurcations (Fig. 5.8 (f)).



(a) original

(b) sum of responses

(c) certainty map

(d) certainty-controlled normalisation of (b)

(e) contour tangent map

(f) spline contours

Figure 5.8: Stages of contour extraction using a photograph of the sculpture "Paolina" (512×512 pixels, from the archive of the Vision group of Pietro Perona at Caltech). The result of certainty-controlled normalisation (d) combines the advantages of (b) and (c), namely, good localisation and separation of linear structures, as well as contrast invariance. The spline representation (f) preserves many of the important contour bifurcations that indicate three-dimensional shape.

5.6.1 The benefits of spline parameter optimisation (tangent updating)

Figure 5.9 illustrates the effect of the tangent optimisation algorithm on the spline contour representation.

When applied directly to the tangent orientations given by the modes of the corresponding (mixture) densities, the "relatability" criterion of Kellman and Shipley (1991) rejects a number of tangent configurations, and consequently a lot of contour segments are not detected. Also, many splines tend to differ from the actual contours, since inexact tangent angles tend to create curved rather than straight splines. Through optimisation, a significant number of tangent angles can be adjusted, in order to yield "relatable" configurations, many of which prove to be consistent with the intensity gradient in the image. As a result, a more complete and accurate contour representation is obtained.

5.6.2 Noise robustness

In another experiment, the performance of the algorithm in the presence of additive Gaussian noise has been investigated. Figure 5.10 shows the result of feature detection and subsequent perceptual grouping for moderate noise.

For a moderate noise level ($\sigma_N < 5\%$, SNR < 26 dB), there are only few false positives in the spline representation, since most erroneous key points form only isolated splines that can easily be identified and removed.

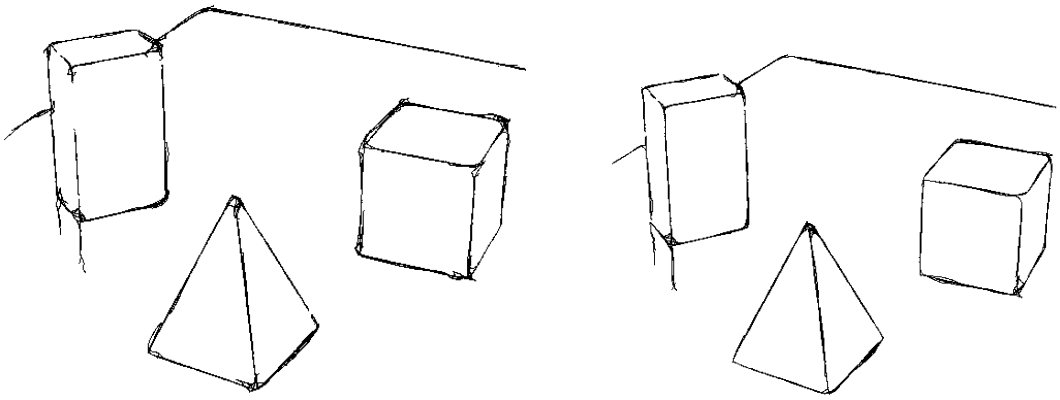
Above a noise value of about 10% (SNR = 20 dB), the density of false positive key points reaches a level where spurious splines begin to form erroneous contour segments of considerable length which could only be eliminated by perceptual organisation of higher order. At this stage curvature consistency would be a vital constraint, since the noise-induced contour segments exhibit frequent, sudden changes in the sign of

curvature, which rarely occur in natural object boundaries and folds. The erroneously *discarded* key points (false negatives) are small in number but, of course, much more obvious, since they lead to gaps in the contour representation.

The number of detected features as a function of the noise level is an indicator for the efficiency of the feature detection in the presence of noise, since the number of additional key points compared to the case without noise approximately equals the number of false positives. Figure 5.12 shows a plot of this relation for the image in Figure 5.10 (a). Though more and more spurious key points appear with increasing noise level, most of them do not fulfill the relatability criterion, and even after angular optimisation, a potential spline connection often lacks consistency with the filter responses. Thus many false positives can be identified and rejected.

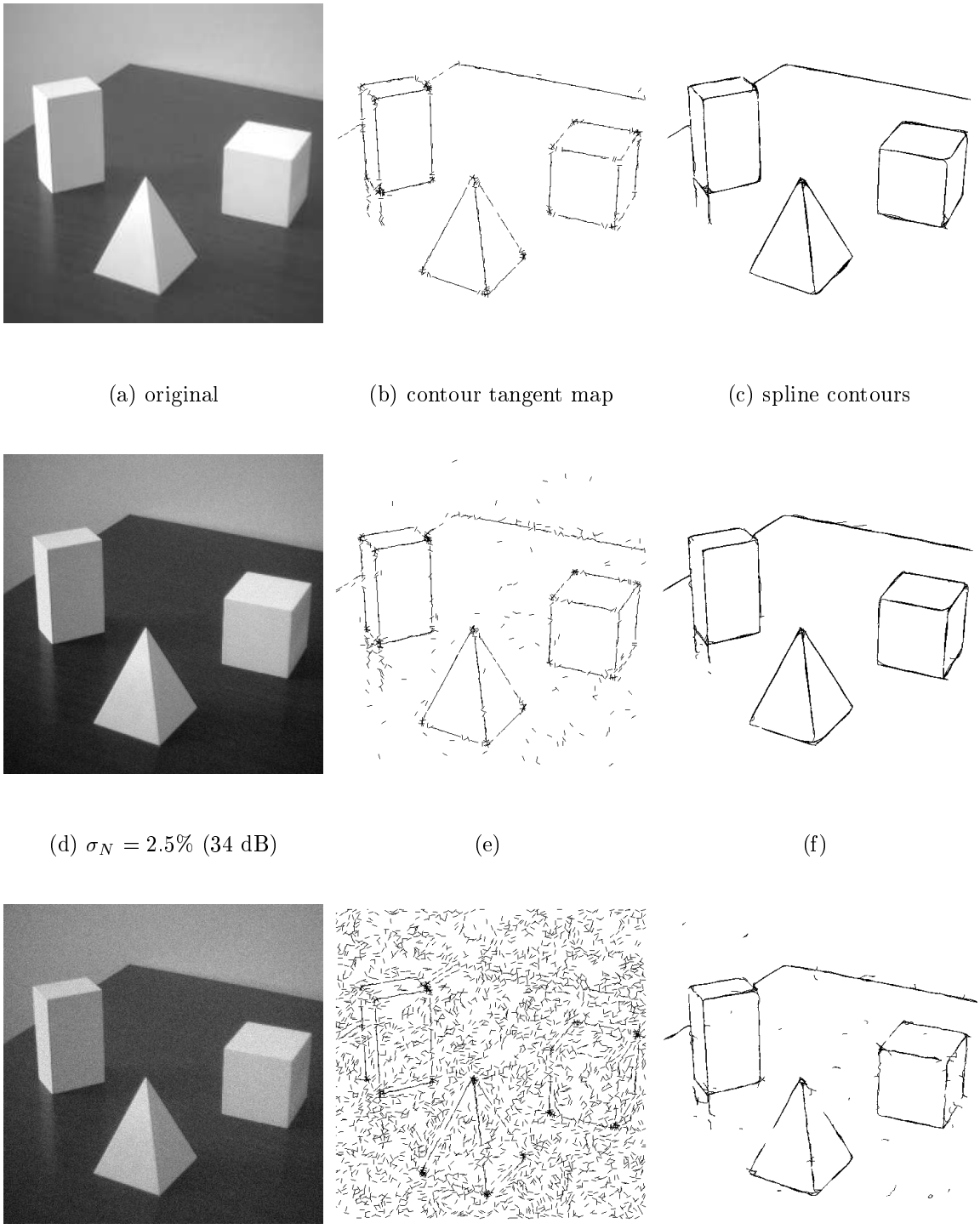
The maximum number of feature points is not merely limited by the number of pixels in the image, but rather by the grid structure imposed on the image during feature selection, whereby the maximum "feature density" becomes less than one key point per pixel. In addition, nearby parallel key point candidates compete with one another for local "supremacy", and many candidates are eliminated. Therefore, the number of key points stagnates above a particular noise level. Since the perceptual grouping algorithm removes a great amount of false positives, their number increases more slowly.

Together, certainty-controlled response normalisation and spline grouping realise a trade-off between noise suppression and amplification of weak responses in order to preserve key points on edges with low contrast. As a result, the algorithm is useful for typical optical images but not for images with extreme noise levels, such as from synthetic aperture radar (SAR) or ultra sound sources.



(a) (b)

Figure 5.9: The effect of tangent optimisation. In (a), the "relatability" criterion of Kellman and Shipley (1991) is applied directly to the tangent orientations given by the modes of the corresponding mixture densities. After tangent optimisation, a higher degree of connectivity is reached and the spline are closer to the actual object contours, since many tangent angles have been adjusted in order to yield "relatable" tangent configurations that prove to be consistent with the intensity gradient in the image.

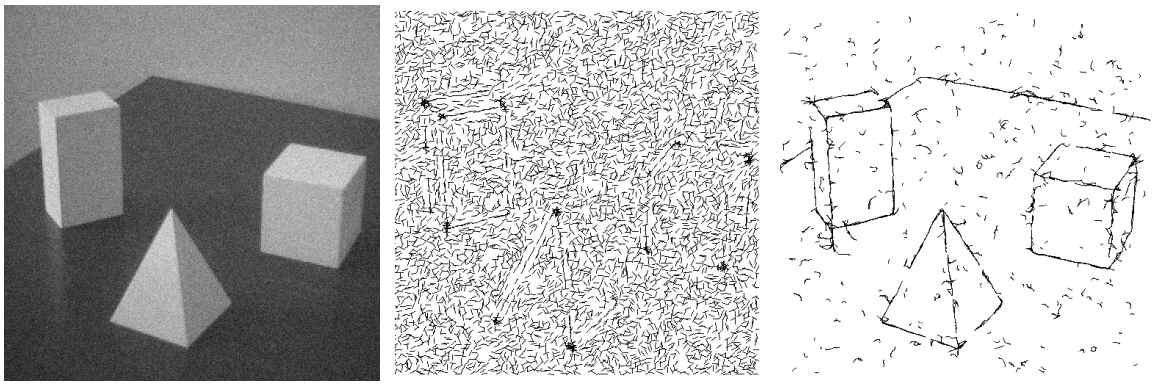


(a) original (b) contour tangent map (c) spline contours

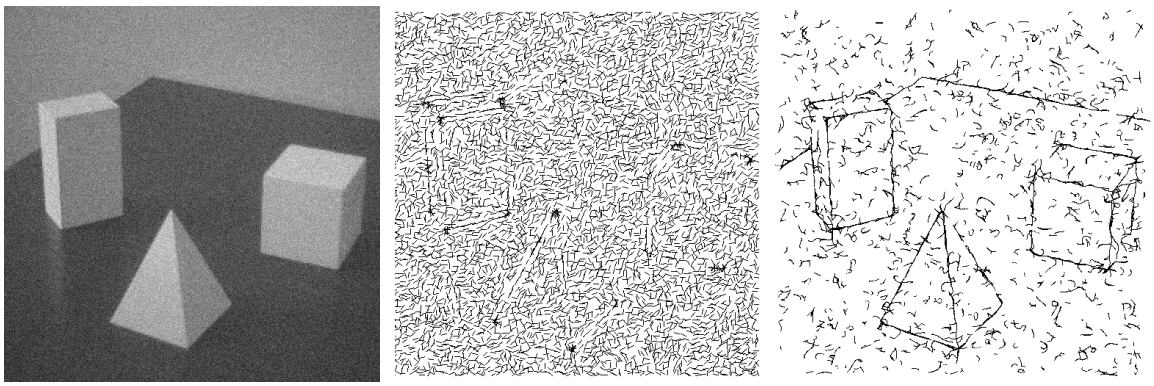
(d) $\sigma_N = 2.5\%$ (34 dB) (e) (f)

(g) $\sigma_N = 5\%$ (26.0 dB) (h) (i)

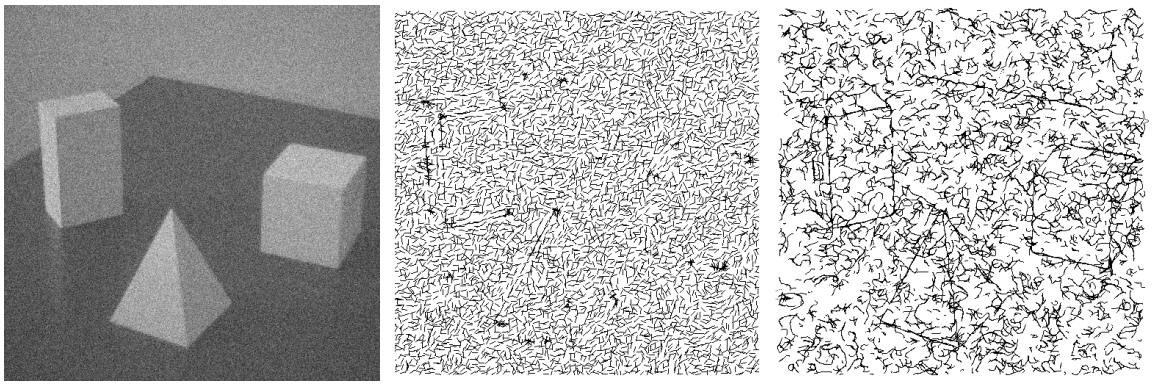
Figure 5.10: An image with different amounts of additive Gaussian noise. Note the relatively small number of false positives in the spline representation.



(a) $\sigma_N = 7.5\%$ (22.5 dB) (b) (c)



(d) $\sigma_N = 10\%$ (20 dB) (e) (f)



(g) $\sigma_N = 15\%$ (16.5 dB) (h) (i)

Figure 5.11: Continuation of Fig. 5.10 with higher noise levels. Above a noise value of about 10% ($SNR = 20$ dB), the density of false positive key points reaches a level where chains of spurious splines begin to form erroneous contour segments of considerable length, which could only be eliminated by perceptual organisation of higher order.

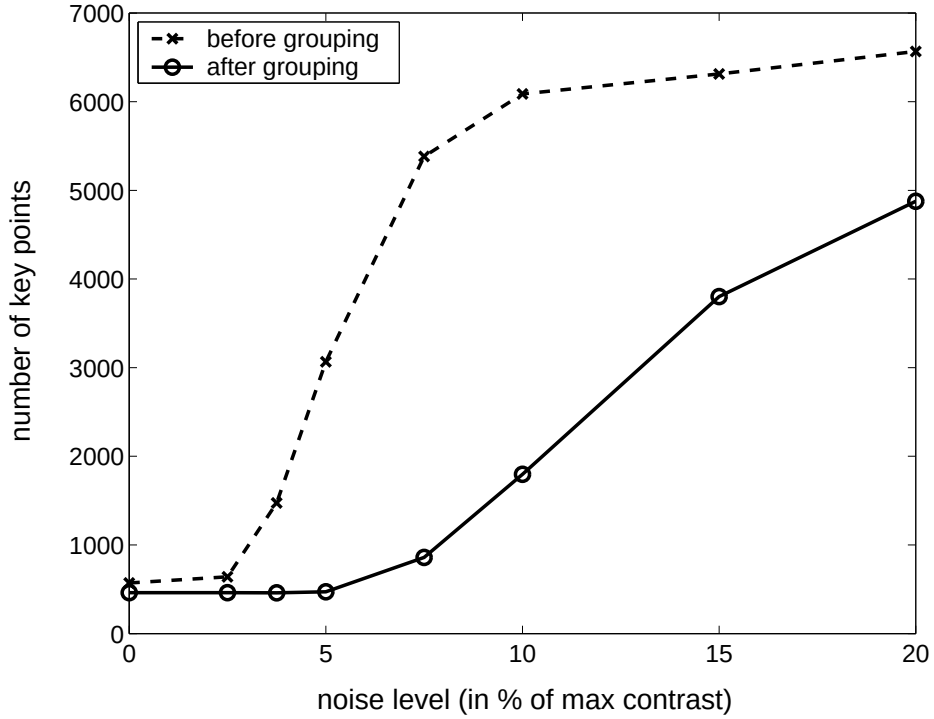


Figure 5.12: The number of key points detected in Figure 5.10 (a) as a function of the noise level (standard deviation in % of maximum contrast) before (dashed curve) and after perceptual organisation (solid curve). Though the number of erroneous key points grows rapidly with increasing noise level, the grouping algorithm is able to eliminate a substantial number of spurious contour segments, and, consequently, the number of remaining false positives increases much more slowly. The noise levels at the data points correspond to Fig. 5.10 and Fig. 5.11. The numbers saturate when an equilibrium is reached between noise-induced key point candidates and competitive feature elimination.

5.7 Conclusions

This chapter has introduced a perceptual grouping framework based on spline interpolation between contour key-points, obtained from population decoding and feature selection through competition among neighbouring edge segments.

From the distributed coding viewpoint, perceptual grouping is interpreted as combining local population codes, accomplished by computing joint densities from the individual probability densities of tangent orientation decoded from the filter ensembles. Thus a probabilistic description of the most suitable interpolating spline arises quite naturally in terms of the likelihood of a pair of key-point tangent angles. Together with an additional bending constraint, the initial tangent angles, given by the modes of the local orientation pdfs, are updated by an MLE-based optimisation algorithm, in order to improve mutual consistency and to overcome errors in the local orientation estimates induced by noise. During tangent optimisation, the explicit representation of local certainty is essential, since the degree of flexibility permitted during the updating of an initial tangent measurement is determined by the concentration parameter of the corresponding pdf, i.e., effectively by its certainty.

The probabilistic population decoding procedure attempts to match a local response model to the actual ensemble activities of the filter bank at every pixel. The ratio of matches and mismatches can help define the consistency between a given spline and the image data (Gabor transform), regardless of the absolute intensity of filter responses. Together with local certainty and the degree of agreement between measured local orientation and the tangent orientation of the spline, a *contrast independent* spline consistency measure has been introduced which tends to create only few false positive contour segments in the presence of moderate image noise, while preserving edges with low contrast.

The spline representation is able to capture perceptually important contour bifur-

cations, such as those arising in regions of self occlusion in folds, which provide vital clues about the 3D shape of the object under consideration. Bifurcations tend to be misrepresented by conventional edge detectors like Canny's (1986), since the responses of linear filters at such points are usually smaller than in their neighbourhood, and tangent discontinuities are often smoothed out.

The fact that, unlike B-splines, Hermite splines are not invariant under affine transformations may be considered a disadvantage. However, the spline representation described in this chapter is to be understood as a *first* grouping procedure, which would require subsequent steps to further organise the short quadratic splines into larger contour segments. It is at this stage that invariance issues would become relevant, but that is beyond the scope of this thesis.

Chapter 6

Conclusions and Outlook

This thesis has presented a biologically motivated approach to edge detection with oriented filters. The novel contribution is the development and application of population coding techniques for the purpose of contour detection in machine vision, inspired by the computational neuroscience literature. This has been accomplished by reconsidering the well-established analogy between Gabor filters and the orientation sensitive "simple cells" in the hypercolumns of mammalian primary visual cortex. Regarding a bank of Gabor filters as an *ensemble* that encodes local edge orientation in a distributed fashion has suggested methods of *de*coding the orientation variable similar to the strategies developed in experimental neurobiology and computational neuroscience, such as the population vector and more refined probabilistic methods.

6.1 Low-level Processing

In particular, the population coding analogy has resulted in the development of a unified probabilistic framework for the detection of one- as well as two-dimensional edge structure. Based on the orientation tuning function of the filters and the assumption that local edge orientation in the presence of noise follows a von Mises distribution,

normal edges and locations with multiple orientations can be modelled by a von Mises mixture distribution, the parameters of which are estimated via a version of the expectation maximisation (EM) algorithm.

One characteristic element of the population coding paradigm is the notion of certainty, which is equivalent to the degree of concentration of activity within the filter ensemble. Once a probability density of orientation is available, certainty can be characterised in terms of its angular variance or its entropy.

It has been demonstrated that an edge detector based on population coding can benefit from utilising the *additional information* provided by the certainty measure. Checking for consistency in the first and second moments of ensemble activities, i.e., the local orientation measurement itself and its variance (or a function thereof), provides a better means of identifying perceptually relevant local features and segregating them from noise.

Localisation of contour features and their characterisation in terms of tangent orientation are very dissimilar tasks that require different processing strategies. While the processing for population coding is kept strictly linear, in order to assure linear superposition of filter responses in the presence of multiple orientations, the feature selection involves non-linear interactions between certainty and response magnitude, in the form of multiplication of the two quantities or a certainty-controlled normalisation within a local neighbourhood.

6.2 Perceptual Grouping

Finally, the information from local ensembles has been combined in order to obtain contour segments. Computing joint probability densities of tangent orientations from key point pairs quite naturally leads to a probabilistic description of interpolating splines as an early form of perceptual organisation. With an additional bending constraint im-

posed on the connecting spline, an optimisation procedure determines the most likely spline, thus smoothing the "jitter" in the initial orientation measurements introduced by image noise.

Since geometric criteria for the association of edge segments are well-established, for example in the form of edge "relatability" (Kellman and Shipley, 1991) or similar concepts such as the "compatibility" relation (Parent and Zucker, 1989), the essential conceptual difference of the grouping algorithm presented in this thesis, compared to these and other approaches, lies in the role of certainty as a *constraint* during the updating of orientation estimates. Moreover, by means of certainty and tangent consistency, it has been possible to define the consistency between spline contour model and the actual intensity structure in the image, *independent* of contour contrast. At the grouping level, the algorithm again benefits from the information provided by local certainty, which characterises the success of matching a response template (i.e., the von Mises mixture model) in each point on a spline. Perceptual "saliency" of a spline is expressed in terms of the number of successful matches compared to the total number of points on the discretised curve.

The fact that a distributed representation of local features can be beneficial for feature grouping justifies linking two seemingly unrelated subjects, population coding in neurobiology and perceptual organisation in machine vision.

6.3 Future Work

Possible extensions of the work of the dissertation, at the feature extraction level, could include the introduction of additional aspects of biological visual processing, such as a Mexican-hat-type interaction within the filter bank. Due to mutual inhibition and excitation, filters in such an intrinsically connected ensemble would have sharper orientation tuning and, consequently, higher angular resolution of the activity profile,

which could be beneficial for the detection of corner points and junctions.

6.3.1 Alternative techniques of population coding

Since the orientation tuning function of Gabor filters is not independent of spatial frequency, the orientation tuning width depends on a free parameter, κ_0, though its value is quite stable across different images, unless contours exhibit a strongly varying degree of defocus. One could eliminate the scale-dependent parameter κ_0 by simultaneously decoding orientation and spatial frequency in a two-dimensional population code, a strategy somewhat similar to the model proposed by Zohary (1992). Such a method would provide an estimate of local characteristic scale. However, computations would become more elaborate, involving larger filter banks.

As an alternative to the mixture model underlying the work of the thesis, a non-parametric approach could be chosen; for example, using the Fourier-transform-based decoding scheme developed by Wilson (Wilson and Lüdtke, 2000).

Perhaps strong curvature could be represented more specifically within such a framework, since the von Mises model cannot accurately describe the flattened response profiles caused by edges with a radius of curvature similar to the size of the filter mask. The degree of deviation of the extracted angular pdf from the von Mises density might then be calculated in terms of the *kurtosis*, yielding a coarse measure of curvature.

6.3.2 Extensions of the grouping algorithm

The perceptual organisation algorithm described in the previous chapter is only the first stage of local feature grouping. The relatively short contour segments represented by quadratic splines could be merged into larger contours, perhaps by using a B-spline model. During such further organisation, the number of control points might be

reduced, leading to a more economical representation resembling that in the approach of Cham and Cipolla (1996), who developed a spline contour framework based on a minimum description length criterion.

Furthermore, the spline interpolation method of Chapter 5 provides a graph representation of the object outlines that could be used for further high-level processing. For instance, graph-theoretic operations, such as a "normalised cut" (Shi and Malik, 1997), could be performed on the spline graph to perform segmentation.

Appendix A

Information Theoretic Measures for the von Mises Distribution

In this chapter some important information-theoretic quantities shall be derived for the case of a single angular random variable with a von Mises distribution

$$p(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \bar{\theta})}\,, \tag{A.1}$$

as they are not given in standard textbooks.

A.1 Fisher Information

The Fisher Information $\mathcal{I}$ for a single variable is defined as (Frieden, 1998)

$$\mathcal{I} = \int_0^{2\pi} \left(\frac{dp(\theta)}{d\theta} \right)^2 \Big/ p(\theta)\, d\theta\,, \tag{A.2}$$

which yields for the von Mises distribution:

$$\mathcal{I}(\kappa) = \frac{\kappa^2}{2\pi I_0(\kappa)} \int_0^{2\pi} \left(-e^{\kappa\cos(\theta-\bar{\theta})}\sin(\theta-\bar{\theta})\right)^2 \Big/ e^{\kappa\cos(\theta-\bar{\theta})}\, d\theta$$

$$= \frac{\kappa^2}{2\pi I_0(\kappa)} \int_0^{2\pi} e^{\kappa\cos(\theta-\bar{\theta})}\sin^2(\theta-\bar{\theta})\, d\theta$$

with integral 3.915 (4.) from (Gradstein and Ryshik, 1981)

$$= \frac{\kappa^2}{2\pi I_0(\kappa)} \sqrt{\pi}\,\tfrac{2}{\kappa}\,\Gamma(1\tfrac{1}{2})I_1(\kappa)$$

$$= \kappa\,\frac{I_1(\kappa)}{I_0(\kappa)}\,\frac{\Gamma(1\frac{1}{2})}{\sqrt{\pi}}\,, \qquad \text{where } \Gamma \text{ is the gamma-function and } \Gamma(1\tfrac{1}{2}) = \sqrt{\pi}/2$$

$$\Longrightarrow \mathcal{I}(\kappa) = \frac{1}{2}\,\frac{I_1(\kappa)}{I_0(\kappa)}\,\kappa$$

□

A.2 Shannon Entropy

The Shannon entropy is defined as (Blahut, 1987)

$$h = -\int_0^{2\pi} p(\theta)\ln p(\theta)\, d\theta\,.$$

Using integrals given in Mardia (1972) or Abramowitz & Stegun (1970) , this yields for the von Mises distribution:

$$
\begin{aligned}
h(\kappa) &= -\frac{1}{2\pi I_0(\kappa)} \int_0^{2\pi} e^{\kappa \cos(\theta-\bar{\theta})} \left\{ -\ln[2\pi I_0(\kappa)] + \kappa \cos(\theta - \bar{\theta}) \right\} \, d\theta \\
&= \frac{\ln[2\pi I_0(\kappa)]}{2\pi I_0(\kappa)} \underbrace{\int_0^{2\pi} e^{\kappa \cos(\theta-\bar{\theta})} \, d\theta}_{2\pi I_0(\kappa)} - \frac{\kappa}{2\pi I_0(\kappa)} \underbrace{\int_0^{2\pi} e^{\kappa \cos(\theta-\bar{\theta})} \cos(\theta - \bar{\theta}) \, d\theta}_{2\pi I_1(\kappa)} \\
\\
\implies h(\kappa) &= \ln[2\pi I_0(\kappa)] - \frac{I_1(\kappa)}{I_0(\kappa)} \kappa \, .
\end{aligned}
$$

□

Bibliography

Abramowitz, M. and Stegun, I. (1970). *Handbook of Mathematical Functions*. Dover Publications Inc., New York.

Adelson, E. and Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2):284–299.

Alter, T. D. and Basri, R. (1998). Extracting salient curves from images: An analysis of the saliency network. *Int. Journal of Computer Vision*, 27(1):51–69.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61:183–193.

August, J. and Zucker, S. (2000). The curve indicator random field: Curve organization via edge correlation. In Boyer, K. and Sarkar, S., editors, *Perceptual Organization for Artificial Vision Systems*, pages 265–288. Kluwer Academic, Boston.

Baldi, P. and Heiligenberg, W. (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biological Cybernetics*, 59:313–318.

Barlow (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.

Blahut, R. E. (1987). *Principles And Practice of Information Theory.* Addison-Wesley Publishing Company.

Buse, R., Liu, Z., and Caelli, T. (1996). Using Gabor filters to measure the physical properties of lines. *Pattern Recognition*, 29(4):615–625.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transact. on Patt. Rec. and Machine Intell.*, 8(6):679–700.

Cham, T.-J. and Cipolla, R. (1996). Automated B-spline curve representation with MDL-based active contours. In *Proceedings of the British Machine Vision Conference.*

Cross, A. and Hancock, E. (1999). Extracting curvilinear features from millimetre radar data. In *Proceedings of the 7th Int. Conf. on Image Processing and its Applications*, pages 537–541.

Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 2:1160–1169.

Deco, G. and Obdradovic, D. (1996). *An Information-Theoretic Approach to Neural Computing.* Springer Verlag.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data. *Journal of the Royal Statistical Society, B*, 39:1–38.

Deneve, S., Latham, P., and Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, 2(8):740–745.

Deneve, S., Latham, P., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4(8):826–831.

Dobbins, A., Zucker, S., and Cynader, M. (1987). End-stopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138):438–441.

Dolan, J. and Riseman, E. (1992). Computing curvilinear structure by token-based grouping. In *Proc. of the IEEE Conf. on Comp. Vision and Pattern Recognition*.

Elder, J. and Zucker, S. (1996). Computing contour closure. In *Proceedings of the European Conference on Computer Vision*, volume Lecture Notes in Computer Science 1064, pages 399–. Springer Verlag.

Ellis, W. D. (1967). *A Source Book of Gestalt Psychology*. Humanities Press, New York.

Farin, G. (1988). *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide*. Academic Press, San Diego, CA.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, 4(12):2379–2394.

Frieden, R. (1998). *Physics From Fisher Information: A Unification*. Cambridge University Press.

Gavrila, D. (1996). Hermite deformable contours. Technical Report CS-TR-3610, Center for Automation Research, University of Maryland. also ICPR'96.

Georgopoulos, A., Schwarz, A., and Kettner, R. (1986). Neural population coding of movement direction. *Science*, 233:1416–1419.

Gilbert, C. and Wiesel, T. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, 30(11):1689–1701.

Gradstein, I. and Ryshik, I. (1981). *Tables of Series , Products and Integrals*. Verlag Harry Deutsch.

Granlund, G. and Knutsson, H. (1995). *Signal Processing in Computer Vision*. Kluwer Academic Publishers.

Grosof, D. H., Shapley, R. M., and Hawken, M. J. (1993). Macaque V1 neurons can signal 'illusory' contours. *Nature*, 365:550–552.

Guy, G. and Medioni, G. (1996). Inferring global perceptual contours from local features. *Int. J. Comp. Vis.*, 20(1/2):113–133.

Hancock, E. and Kittler, J. (1990). Discrete relaxation. *Pattern Recognition*, 23(7):711–733.

Heeger, D., Simoncelli, E., and Movshon, J. (1996). Computational models of cortical processing. *Proceedings of the National Academy of Sciences, USA*, 93:623–627.

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. *Vision Research*, 32(5):963–981.

Heitger, F. and von der Heydt, R. (1993). A computational model of neural contour processing: Figure-ground segregation and illusory contours. In *Proc. of the Fourth Int. Conf. on Comp. Vis.*

Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., and Kübler, O. (1998). Simulation of neural contour mechanisms: Representing anomalous contours. *Image and Vision Computing*, 16:407–421.

Hess, R., Dakin, S., and Field, D. (1998). The role of "contrast enhancement" in the detection and appearance of visual contrast. *Vision Research*, 38(6):783–787.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154.

Itti, L., Braun, J., Lee, D., and Koch, C. (1998). A model of early visual processing. In *Adv. in neur. Inf. Proc.*, volume 10, pages 173–179.

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

Kanizsa, G. (1979). *Organization in Vision*. Praeger, New York.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *Int. Journal of Computer Vision*, 1:321–332.

Kellman, P. and Shipley, T. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23:141–221.

Koendrink, J. and van Doorn, A. (1982). The shape of smooth objects and the way contours end. *Perception*, 11:129–137.

Krüger, N. (1998). Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*, 8:117–129.

Krüger, N. and Lüdtke, N. (1998). Orassyll: Object recognition with autonomously learned and sparse symbolic representations based on local line detectors. In *Proceedings of the British Machine Vision Conference*.

Krüger, N. and Peters, G. (1997). Object recognition with banana wavelets. In *Proceedings of the ESANN*.

Lades, M., Vorbrüggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311.

Lehky, S. and Sejnowski, T. (1990). Neural model of stereo-acuity and depth interpolation based on a distributed representation of stereo disparity. *Journal of Neuroscience*, 10:2281–2299.

Lehky, S. and Sejnowski, T. (1998). Seeing white: Qualia in the context of population decoding. *Neural Computation*, 11:1261–1280.

Leite, J. and Hancock, E. (1997). Iterative curve organization. *Pattern Recognition Letters*, 18:143–155.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10:903–940.

Lowe, D. (1985). *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Boston, MA.

Lüdtke, N. and Hancock, E. (2001). Gabor population codes for orientation selection. In Musio, C., editor, *Vision, The Approach of Biophysics and Neurosciences. Proceedings of the International School of Biophysics*, volume 11 of *Series on Biophysics and Biocybernetics*, pages 437–440, Casamicciola, Naples, Italy. Istituto Italiana Per Gli Studi Filosofici, World Scientific.

Lüdtke, N., Wilson, R., and Hancock, E. (2000a). Population codes for orientation estimation. In *Proceedings of the 15th International Conference on Pattern Recognition*.

Lüdtke, N., Wilson, R., and Hancock, E. (2000b). Tangent fields from population coding. In *Proceedings of the IEEE International Workshop on Biologically Motivated Computer Vision (BMCV2000)*, volume 1811 of *Lecture Notes in Computer Science*, pages 584–593.

Lüdtke, N., Wilson, R., and Hancock, E. (2002a). Population coding of multiple edge orientation. In Dorronsoro, J., editor, *Proceedings of the International Conference on Artificial Neural Networks*, volume 2415 of *Lecture Notes in Computer Science*, pages 1261–1267.

Lüdtke, N., Wilson, R., and Hancock, E. (2002b). Probabilistic population coding of multiple edge orientation. In *Proceedings of the International Conference on Image Processing*, volume 2, pages 865–868. IEEE.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70:1297–1300.

Mardia, K. (1972). *Statistics of Directional Data*. Academic Press, London and New York.

Morrone, M. and Owens, R. (1987). Feature detection from local energy. *Pattern Recognition Letters*, 6:303–313.

Oram, M., Földiàk, P., Perrett, D., and Sengpiel, F. (1998). The 'ideal homunculus': Decoding neural population signals. *Trends in Neuroscience*, 21(6):259–265.

Papoulis, A. (2002). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, fourth edition.

Paradiso, M. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics*, 58:35–49.

Parent, P. and Zucker, S. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:823–839.

Perez de la Blanca, N., Fdez-Valdivia, J., and Garcia, J. (1993). Characterizing planar outlines. *Pattern Recognition Letters*, 14:489–497.

Perona, P. (1992). Steerable-scalable kernels for edge detection and junction analysis. *Image and Vision Computing*, 10(10):663–672.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., and Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 5:580–584.

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. *Nature Neuroscience Reviews*, 1:125–132.

Pouget, A. and Zhang, K. (97). Statistically efficient estimation using cortical lateral connections. In *Advances in Neural Information Processing*, volume 9, pages 97–

103.

Press, W., Flannery, B., Teukolski, S., and Vetterling, W. (1989). *Numerical Recipes*. Cambridge University Press.

Rumelhart, D., McClelland, J., and PDP Research Group (1986). *Parallel Distributed Processing*. MIT Press, Cambridge, MA.

Salinas, E. and Abbott, L. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience*, 1:89–107.

Sanger, D. (1994). Theoretical considerations for the analysis of population coding in motor cortex. *Neural Computation*, 6:29–37.

Sanger, D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology*, 76(4):2790–2793.

Sha'ashua, A. and Ullman, S. (1988). Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the 2nd ICCV*, pages 321–327. Also Weizmann Institute of Science Report CS88-18, (October 1988).

Shi, J. and Malik, J. (1997). Normalized cuts and image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shustorovich, A. (1994). Scale specific and robust edge/line encoding with linear combinations of Gabor wavelets. *Pattern Recognition*, 27(5):713–725.

Snippe, H. (1996). Parameter extraction form population codes: A critical assessment. *Neural Computation*, 8:511–529.

Stemmler, M., Usher, M., and Niebuhr, E. (1995). Lateral interactions in primary visual cortex: A model bridging physiology and psychophysics. *Science*, 269:1877–1880.

Swindale, N. (1998). Orientation tuning curves: Empirical description and estimation of parameters. *Biol. Cybern.*, 78:45–56.

Tanaka, K. (1996). Representation of visual features of objects in the inferotemporal cortex. *Neural Networks*, 9(8):1459–1475.

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

von der Heydt, R. and Peterhans, E. (1989a). Mechanisms of contour perception in monkey visual cortex. I. lines of pattern discontinuity. *J. of Neuroscience*, 9(5):1731–1748.

von der Heydt, R. and Peterhans, E. (1989b). Mechanisms of contour perception in monkey visual cortex. II. bridging gaps. *Journal of Neuroscience*, 9(5):1749–1763.

Wilson, M. and McNaughton, B. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261:1055–1058.

Wilson, R. and Lüdtke, N. (2000). Decoding population codes. In *Proceedings of the 15th International Conference on Patern Recognition*, volume 2, pages 137–140.

Young, M. and Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, 256:1327–1331.

Zemel, R. and Dayan, P. (1997). Combining probabilistic population codes. In *Proc of the Int. Joint Conf. on Artif. Intell. (IJCAI97).*

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

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

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

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

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

Zhang, K. and Sejnoswki, T. (1999). Neural tuning: To sharpen or broaden? *Neural Computation*, 11:75–84.

Zohary, E. (1992). Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biological Cybernetics*, 66:265–272.

Zucker, S., David, C., Dobbins, A., and Iverson, L. (1988). The organization of curve detection: Coarse tangent fields and fine spline coverings. In *Proc. of the 2nd Int. Conf. on Comp. Vision (ICCV'88)*, pages 577–586.

Zucker, S., Dobbins, A., and Iverson, L. (1989). Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1:68–81.