

# Direct computation of shape cues by multi-scale retinotopic processing

Jonas Gårding and Tony Lindeberg

Computational Vision and Active Perception Laboratory (CVAP)\*  
Department of Numerical Analysis and Computing Science  
Royal Institute of Technology, S-100 44 Stockholm, Sweden  
Email: tony@bion.kth.se, jonasg@bion.kth.se

*Submitted. Technical report ISRN KTH/NA/P-93/05-SE..*

## Abstract

This paper addresses the problem of computing cues to the three-dimensional structure of surfaces in the world directly from the local structure of the brightness pattern of either a single monocular image or a binocular image pair.

It is shown that starting from Gaussian derivatives of order up to two at a range of scales in scale-space, local estimates of (i) surface orientation from monocular perspective texture foreshortening, (ii) surface orientation or curvature from monocular texture gradients, and (iii) surface orientation from the binocular disparity gradient can be computed without iteration or search, and by using essentially the same basic mechanism.

The methodology is based on a multi-scale descriptor of image structure called the windowed second moment matrix, which is computed with adaptive selection of both scale levels and spatial positions. Notably, this descriptor comprises two scale parameters, a local scale describing the amount of smoothing used in derivative computations, and an integration scale determining over how large a region in space the statistics of local descriptors is accumulated.

Experimental results for both synthetic and natural images are presented, and the relation with models of biological vision is briefly discussed.

---

\*We would like to thank Jan-Olof Eklundh for continuous support and encouragement, as well as Narendra Ahuja at University of Illinois, and John P. Frisby at University of Sheffield for kindly providing several of the images used in the paper. This work was partially performed under the ESPRIT-BRA project INSIGHT. The support from the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged.

The first author has carried out part of this work while visiting the AIVRU group at University of Sheffield, and he is grateful for their hospitality as well as for the financial support of the Foundation Blanceflor Boncompagni-Ludovisi, née Bildt, and the Swedish Institute.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>A local texture descriptor</b>	<b>3</b>
2.1	The windowed second moment matrix . . . . .	3
2.1.1	Spatial frequency interpretation . . . . .	4
2.1.2	Visualization by ellipses . . . . .	4
2.2	Transformation properties . . . . .	5
2.3	The structure of the second moment descriptor . . . . .	6
<b>3</b>	<b>Representing and selecting scale</b>	<b>7</b>
3.1	Scale selection: Basic principle . . . . .	8
3.2	Proposed method for scale selection . . . . .	9
3.3	Properties of the scale selection method . . . . .	11
3.4	Uniqueness of the window function . . . . .	11
3.5	Scale selection for computation of $\mu_L$ . . . . .	12
3.6	The ellipse representation revisited . . . . .	15
<b>4</b>	<b>Spatial selection and blob detection</b>	<b>16</b>
4.1	Spatial selection: Basic principle . . . . .	16
4.2	Experimental results . . . . .	18
<b>5</b>	<b>Shape from texture</b>	<b>20</b>
5.1	Background . . . . .	20
5.2	Review of image geometry . . . . .	20
5.3	Deriving shape cues from the second moment descriptor . . . . .	22
5.3.1	Shape from foreshortening . . . . .	23
5.3.2	Shape from the area gradient . . . . .	23
5.4	Estimating surface shape and orientation: Basic scheme . . . . .	24
5.5	The texel grouping scale . . . . .	25
5.6	Experimental results . . . . .	25
<b>6</b>	<b>Shape from disparity gradients</b>	<b>29</b>
6.1	Review of stereo geometry . . . . .	30
6.2	Estimating disparity gradients . . . . .	31
6.3	Experimental results . . . . .	32
<b>7</b>	<b>Summary and discussion</b>	<b>34</b>
7.1	Relations to biological vision . . . . .	34
7.2	Further research . . . . .	35
<b>A</b>	<b>Appendix</b>	<b>36</b>
A.1	Transformation property of the second moment matrix . . . . .	36
A.2	Estimating simple distortion gradients . . . . .	37



# 1 Introduction

Virtually all methods for inferring properties of the three-dimensional world from one or more images require an initial stage of retinotopic processing in which the raw image brightness pattern is transformed into some more useful representation. In practical computer vision applications this representation is often tailored for the specific task at hand, but a number of attempts have been made at defining general principles for the structure of a more general-purpose set of low-level operators capable of computing useful representations without any specific prior knowledge of the image structures to be processed.

One such approach, based primarily on theoretical considerations, is the *scale-space representation*, introduced by Witkin (1983) and Koenderink (1984). Perhaps the most important conclusion of this theory is that if the low-level operators are unbiased in the sense that they do not single out particular locations, orientations, or sizes, then the only permissible linear operations are convolutions with Gaussian kernels and their derivatives at various scales.

An alternative approach is to try to emulate the structure and characteristics of primate vision, either for the purpose of gaining a better understanding of it, or simply because the performance of biological vision systems is superior to that of existing computer vision systems. This approach is somewhat hampered by the fact that current knowledge of biological vision is incomplete, so the evaluation of the aptness of any particular model often proves to be quite difficult. Nevertheless, the approach has generated many interesting and useful results; for example, general considerations regarding the information processing requirements of the visual system led Marr (1976) to propose the computation of a *primal sketch* in which low-level features of the brightness pattern, such as bars and blobs, are explicitly represented. Other models, e.g. (Turner 1986; Bergen and Adelson 1988; Malik and Perona 1990), have been based on neurobiological studies of the structure of the receptive fields in the mammalian retina and the visual cortex. These models have been quite successful at predicting human pre-attentive texture discrimination, and have largely replaced the earlier texton theory by Julesz (1981). Interestingly, the theoretical scale-space approach and the more empirical receptive field approach are to a certain extent in agreement; simple receptive fields in the mammalian retina and visual cortex are well described by Gaussian derivatives (Young 1985, 1987; Jones and Palmer 1987a,b), but also by similar models such as Gabor functions.

Retinotopic processing models are often based on considerations of relatively low-level visual tasks, such as feature detection and two-dimensional texture discrimination. One might therefore be led to think that visual tasks concerning three-dimensional interpretations of the environment require a qualitatively different type of information processing, which would have little in common with such basic operations as can be performed by a single cell or processing unit. In this paper, however, we show that at least some visual tasks of this type can be implemented as bottom-up retinotopic processing sequences, without the need for iterations, search, or a priori knowledge.

More specifically, we consider the task of estimating the shape and orientation of three-dimensional surfaces in the scene from (i) perspective distortion of surface

texture observed in a monocular image, and (ii) the gradient of disparity observed in a binocular image pair. We show that this can be achieved using in principle only the following types of operations: (large support) diffusion smoothing, (small support) derivative computations from smoothed brightness data, and (pointwise) non-linear combinations of these derivatives.

The framework is based on the computation of a local (regional) descriptor of the structure of the brightness pattern, referred to as the *windowed second moment matrix*, which describes the local variance of blurred first-order directional Gaussian derivatives. We emphasize and analyze the need for two different scale parameters; a *local scale* parameter describing the amount of smoothing used for suppressing irrelevant fine scale structures when computing pointwise non-linear descriptors of the image brightness pattern, and a second *integration scale* parameter describing the size of the spatial window used for accumulating statistics of the pointwise descriptors.

Thus, the multi-scale nature of image structures is explicitly taken care of, and is built into the representation. We do not attempt to make the representation “complete” in the sense of allowing reconstruction of the original image from the descriptors. On the contrary, we emphasize adaptive *selection* of both scale levels and spatial positions, for the purpose of providing an explicit representation of precisely the information needed by the later stage processes. Moreover, the representation is normalized in such a way that selection of interesting scale levels and spatial positions is achieved simply through detection of local extrema with respect to scale and position of the computed non-linear entities.

Although the proposed methodology is developed with specific visual tasks in mind, the principles behind it are quite general, which leads us to believe that it will prove to be useful for other tasks as well. It should be made clear that we have attempted to model biological vision only in the very general sense of using basic operations similar to those performed by certain simple cells in the visual cortex. More importantly, however, we do believe that the resulting methodology is a significant contribution to the field of machine vision, with regard to applicability, simplicity and robustness.

The presentation is organized as follows. Section 2 provides a formal definition and description of the basic multi-scale image texture descriptor we propose. Section 3 describes scale problems arising in this context. The notions of local scale and integration scale are formalized, and a methodology for automatic selection of the two scale parameters is presented. Section 4 demonstrates how the basic principles for scale selection can be extended to spatial selection, resulting in what can be viewed as a multi-scale blob detection method. These components are then combined in Section 5, which reviews the shape-from-texture problem and demonstrates how estimates of surface shape and orientation can be computed directly from the multi-scale texture descriptor. Section 6 treats the problem of estimating shape from gradients of binocular disparity, and demonstrates that the proposed approach can be successfully applied to this problem as well. Finally, in Section 7 some general conclusions are made, and their implications are discussed.

## 2 A local texture descriptor

The task of computing meaningful texture descriptors is often referred to in the literature as extraction of texture elements or “texels”. Considering the great variability of natural textures, it is not surprising that there is no generally accepted definition of precisely what a texel is. A first and rather obvious requirement on a texel definition is that it must be computable for a large class of natural images, but this still leaves many degrees of freedom.

Here, we will take a functional approach to texel extraction; rather than postulating any particular structure of the texture, we consider the requirements of the higher-level processes that need to use the local texture description. The basic principle of shape-from-texture estimation is to use the observed perspective distortion of the texture pattern to estimate the parameters of the distorting transformation, which in turn allow properties of surface and/or viewing geometry to be inferred. The principle of shape-from-disparity-gradient estimation is analogous, the difference being that it uses the distortion from the right to the left image, rather than the distortion from surface to image. Hence, for both these processes, the texture description must reflect perspective distortion of the texture in a predictable way, so that the parameters of the distorting transformation can be recovered from the texture description.

A great simplification of the problem comes from the observation that for most purposes it is only necessary to recover the *linear* part of the perspective distortion. The analysis behind this observation is given in Sections 5 and 6; for the moment we take it as a given fact.

### 2.1 The windowed second moment matrix

We propose that a texture descriptor expressed in the form of a two-dimensional *second moment matrix* is ideally suited for the purpose of estimating local linear distortion. Such a second moment matrix can be thought of e.g. as a covariance matrix of a two-dimensional random variable, or, with a mechanical analogy, as the moment of inertia of a mass distribution in the plane. It can be graphically represented by an ellipse, and as will be shown, a linear transformation applied to the spatial coordinates affects the ellipse precisely as it would affect a physical ellipse painted on the surface.

Various forms of second moment descriptors have previously been successfully applied to a number of tasks. For estimation of shape from texture, Brown and Shvaytser (1990) used the second moment of the image brightness autocorrelation function to estimate foreshortening, Gårding (1991) used the second moment of the local Fourier spectrum to estimate foreshortening and several texture gradients, and Super and Bovik (1992) used the same moment to estimate relative foreshortening. Second moments of the directional statistics of image contours have been used by Kanatani (1984), Blake and Marinos (1990), and Gårding (1993) for estimation of foreshortening. Moreover, second moment descriptors of brightness gradients have been used by Bigün et al (1991), Rao and Schunk (1991) for analysis of oriented or flow-like texture patterns, and by Förstner and Gülch (1987) as an “interest” operator in the context of junction detection and stereo matching.

Here, we will use a particular type of second moment matrix similar to some of those described in the above cited articles. It is defined as follows. Let  $L : \mathbb{R}^2 \rightarrow \mathbb{R}$  be the image brightness, and let  $\nabla L = (L_x, L_y)^T$  be its gradient. We now define the second moment descriptor<sup>1</sup>  $\mu_L : \mathbb{R}^2 \rightarrow \text{SPSD}(2)$  of  $L$  by

$$\mu_L(q) = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} = E_q \begin{pmatrix} L_x^2 & L_x L_y \\ L_x L_y & L_y^2 \end{pmatrix} = E_q((\nabla L)(\nabla L)^T), \quad (1)$$

where  $E_q$  denotes an averaging operator centered at  $q = (x, y)^T \in \mathbb{R}^2$ .  $\mu_L(q)$  has a number of convenient properties. Clearly, it is invariant to translations, and it can easily be shown that the trace and determinant of  $\mu_L$  are also invariant to rotations. Moreover, uniform rescaling in the spatial domain and affine brightness transformations only affect  $\mu_L$  by a uniform scaling factor.

We define the averaging operator  $E_q$  as the local weighted mean using a symmetric and normalized window function  $w : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Hence, the components  $\mu_{ij}$  of  $\mu_L(x, y)$  can be expressed as

$$\mu_{ij}(x, y) = \iint_{(x', y') \in \mathbb{R}^2} w(x - x', y - y') L_{x_i}(x', y') L_{x_j}(x', y') dx' dy', \quad (2)$$

The invariance properties are preserved provided that  $w$  is rotationally symmetric (see below) and has a nice scaling behaviour. A natural choice of window function is the Gaussian; in fact, it will be shown in Section 3.4 that this is the *only* translationally invariant choice that leads to scale-space behaviour of  $\mu_L$ .

### 2.1.1 Spatial frequency interpretation

$\mu_L$  can also be understood in terms of the spatial frequency distribution of  $L(x, y)$ . Rename temporarily the coordinates  $(x, y)^T$  to  $(x_1, x_2)^T$ , and let  $\Phi_L : \mathbb{R}^2 \rightarrow \mathbb{R}$  be the power spectrum of  $L$ , i.e.,

$$\Phi_L(\omega_1, \omega_2) = \hat{L}(\omega_1, \omega_2) \hat{L}^*(\omega_1, \omega_2),$$

where  $\hat{L} : \mathbb{R}^2 \rightarrow \mathbb{C}$  denotes the Fourier transform of  $L$  and  $\hat{L}^*$  its complex conjugate. Using Plancherel's relation it follows that

$$\iint_{(x_1, x_2) \in \mathbb{R}^2} L_{x_i} L_{x_j} dx_1 dx_2 = \frac{1}{(2\pi)^2} \iint_{(\omega_1, \omega_2) \in \mathbb{R}^2} \omega_i \omega_j \Phi_L(\omega_1, \omega_2) d\omega_1 d\omega_2. \quad (3)$$

Hence, if  $L \in \mathbb{L}_2(\mathbb{R}^2)$ , the inner products of the first derivatives are proportional to the components of the second moment of the power spectrum.

### 2.1.2 Visualization by ellipses

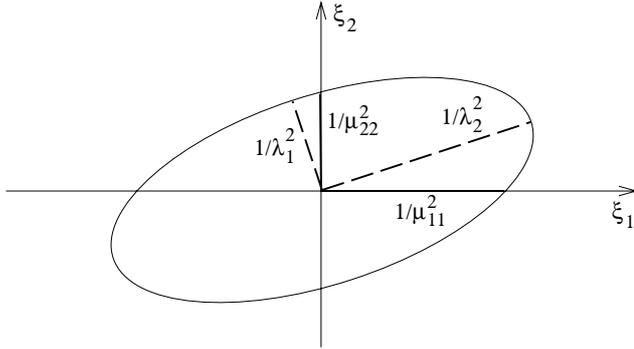
Since the second moment matrix is positive semidefinite, it follows that the equation

$$(\xi - q)^T \mu_L(q) (\xi - q) = 1 \quad (\xi, q \in \mathbb{R}^2) \quad (4)$$

---

<sup>1</sup>The notation  $\text{SPSD}(2)$  stands for the cone of symmetric positive semidefinite  $2 \times 2$  matrices.

defines an ellipse (possibly degenerated to a line) centered at  $q$ . The semi-axes of this ellipse are the square roots of the inverse of the eigenvalues of  $\mu_L(q)$ , while the orientations of the axes give the directions of the corresponding eigenvectors (see Figure 1). It is easily verified that the distance from the center to the perimeter of the ellipse in some direction is equal to the inverse of the average squared magnitude of the directional derivative of  $L(x, y)$  in that direction.



**Figure 1:** The ellipse representation of the second moment matrix  $\mu_L$ . For simplicity, the ellipse is shown centered at the origin of the coordinate system.

## 2.2 Transformation properties

As mentioned in the beginning of this section, the (linear) transformation properties of the local texture descriptor are crucial to the higher-level processes (shape-from-texture and shape-from-disparity-gradients) that are going to operate on the description. Because these processes attempt to recover the parameters of the transformation from the properties of the texture descriptors, the descriptors must be affected in a predictable way by a linear transformation  $B : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , representing e.g. the linearized perspective mapping from the surface to the image in the shape-from-texture case, or the linearized projective mapping from the left to the right image in the shape-from-disparity-gradient case.

For the windowed second moment matrix the relation is straightforward. Given a brightness pattern  $L$ , let  $R : \mathbb{R}^2 \rightarrow \mathbb{R}$  represent the brightness pattern subjected to an invertible linear transformation of the spatial coordinates  $\eta = B\xi$ , i.e.,

$$L(\xi) = R(B\xi) \quad (5)$$

where  $\xi, \eta \in \mathbb{R}^2$ . Moreover, let  $\mu_R(p) \in \text{SPSD}(2)$  be the local second moment of  $R$  at the point  $p = Bq$  computed with respect to the “backprojected” normalized window function  $w' : \mathbb{R}^2 \rightarrow \mathbb{R}$  defined by

$$w'(\eta - p) = (\det B)^{-1} w(B^{-1}(\eta - p)) = (\det B)^{-1} w(\xi - q). \quad (6)$$

It is then straightforward to show that (see Appendix A.1)

$$\mu_L(q) = B^T \mu_R(p) B. \quad (7)$$

In the rest of this section, the arguments  $p$  and  $q$  to  $\mu_L$  and  $\mu_R$  will be dropped to simplify the notation.

It is easily verified that (7) also describes the effect of the coordinate transformation  $B$  to the ellipse (4) representing  $\mu_L(q)$ . Hence, it is justifiable to think of  $\mu_L$  as analogous to an ellipse that is “painted” on the surface. This analogy often provides sufficient intuition to directly predict the behaviour of  $\mu_L$  in various situations.

If  $\mu_L$  and  $\mu_R$  are known, then the linear transformation  $B$  is clearly constrained by (7). However, it is not determined uniquely, since  $\mu_L$  and  $\mu_R$  are symmetric and hence only contain three independent components, whereas  $B$  contains four unknown parameters. It can be shown (Gårding 1991) that the general solution to  $\mu_L = B^T \mu_R B$  is

$$B = \mu_R^{-1/2} W^T \mu_L^{1/2} \quad (8)$$

where  $W$  is an arbitrary orthogonal matrix, and the notation  $\mu^{1/2}$  indicates the unique positive definite symmetric solution to the equation  $X^2 = \mu$ .

Fortunately, in the applications to shape estimation from texture and disparity gradients considered in this paper, it turns out that the ambiguity represented by the rotation matrix  $W$  is of no consequence.

### 2.3 The structure of the second moment descriptor

In this section we will take a closer look at the structure of  $\mu_L(q)$ , and define a number of derived entities that will turn out to be useful later on.

For any two-dimensional second moment matrix  $\mu$ , the following entities can be defined from its components  $\mu_{ij}$ :

$$P = \mu_{11} + \mu_{22}, \quad C = \mu_{11} - \mu_{22}, \quad S = 2 \mu_{12}. \quad (9)$$

Applied to  $\mu_L$  (with the argument  $q$  dropped), these definitions can be rewritten:

$$P = E_q(L_x^2 + L_y^2), \quad C = E_q(L_x^2 - L_y^2), \quad S = 2 E_q(L_x L_y). \quad (10)$$

The first descriptor  $P : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a natural measure of the strength of operator response; it is the average of the square of the gradient magnitude in a region around  $q$ . The two other entities  $C, S : \mathbb{R}^2 \rightarrow \mathbb{R}$  contain directional information, the magnitude of which is

$$Q = \sqrt{C^2 + S^2}. \quad (11)$$

We also define the normalized entities

$$\tilde{C} = C/P, \quad \tilde{S} = S/P, \quad \tilde{Q} = Q/P. \quad (12)$$

It can easily be shown that  $\tilde{Q} \in [0, 1]$ ; it holds that  $\tilde{Q} = 0$  if and only if  $E_q(L_x^2) = E_q(L_y^2)$  and  $E_q(L_x L_y) = 0$ , while  $\tilde{Q} = 1$  if and only if  $E_q(L_x L_y) = E_q(L_x)E_q(L_y)$ .  $\tilde{Q}$  is a natural measure of the *anisotropy* of  $\mu_L(q)$ ; in terms of the ellipse representation,  $\tilde{Q} = 0$  corresponds to a circle, and  $\tilde{Q} = 1$  to a line. For example, a rotationally symmetric brightness pattern has  $\tilde{Q} = 0$ , while a translationally symmetric pattern<sup>2</sup>

<sup>2</sup>A (two-dimensional) brightness pattern  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  which can be written  $f(x, y) = h(ax + by)$  for some one-dimensional function  $h : \mathbb{R} \rightarrow \mathbb{R}$  and some (scalar) constants  $a$  and  $b$ .

has  $\tilde{Q} = 1$ . Rotational symmetry is, however, not necessary in order to obtain  $\tilde{Q} = 0$ . For example, any pattern with  $N \geq 2$  uniformly distributed dominant (unsigned) directions also satisfies  $\tilde{Q} = 0$ . A second moment matrix with  $\tilde{Q} = 0$  will be referred to as *weakly isotropic*.

$Q$  and  $P$  are independent of the coordinate system if the window function  $w$  is rotationally symmetric, and they allow the differential invariants of  $\mu_L$  to be succinctly expressed as follows:

$$\begin{aligned} \text{trace } \mu_L &= P, \\ \det \mu_L &= \frac{1}{4}(P^2 - Q^2) = \frac{1}{4}P^2(1 - \tilde{Q}^2), \\ \lambda_{1,2} &= \frac{1}{2}(P \pm Q) = \frac{1}{2}P(1 \pm \tilde{Q}), \end{aligned} \tag{13}$$

where  $\lambda_1 \geq \lambda_2$  are the eigenvalues of  $\mu_L$ .

The normalized components  $(\tilde{C}, \tilde{S})^T$  can also be understood as representing the local statistics of unsigned gradient directions. A standard technique (Mardia 1972) for computing statistics of unsigned directions in  $\mathbb{R}^2$  is to map a direction angle  $\alpha$  to the point  $(\cos 2\alpha, \sin 2\alpha)^T$  on the unit circle. This mapping has the desired property that  $\alpha$  and  $-\alpha$  are mapped to the same point. Using this representation, map each gradient vector  $(L_x, L_y)^T = \rho(\cos \alpha, \sin \alpha)^T$  to the point  $(\cos 2\alpha, \sin 2\alpha)^T$ , and give it a “mass” proportional to the squared gradient magnitude  $\rho^2$  multiplied by the window function. It is then easily shown that the center of mass of this distribution is given by  $(\tilde{C}, \tilde{S})^T$ . Hence, the average unsigned gradient direction is  $\arg(\tilde{C}, \tilde{S})/2$ , which is also the direction of the eigenvector corresponding to the largest eigenvalue of  $\mu_L$ ; see (Lindeberg and Gårding 1993) for more details.

### 3 Representing and selecting scale

An intrinsic property of objects in the world and details in images is that they only exist as meaningful entities over certain ranges of scale. This issue is of crucial importance when using perspective distortion of the brightness pattern to derive shape cues; size variations of image structures can occur both because a surface texture contains structures at different scales, and because of the perspective distance effects in the image formation process. Analysing image structures at wrong scales often leads to meaningless results. Concerning the computation of the windowed second moment matrix (or, indeed any other non-trivial texture descriptor which involves integration of statistics of pointwise properties over finite-sized local image neighborhoods) there are two fundamental scale problems, which manifest themselves as follows.

First, the image statistics must be collected from a region large enough to be representative of the texture. Yet, the region must not be so large that the local linear approximation of the perspective mapping becomes invalid. For example, for an ideal texture consisting of isolated blobs, a lower limit for the extent of the integration region is determined by the size of the individual blobs, while an upper limit may be given by the curvature of the surface or interference with other nearby surface

patches. This scale controlling the *window function* is referred to as *integration scale* (denoted  $s$ ).

Second, the image statistics must be based on descriptors computed at proper scales, so that noise and “irrelevant” image structures can be suppressed. The descriptor considered in this paper is based on first order spatial derivatives of the image brightness, and it is obvious that useful results hardly can be expected if the derivatives are computed directly from unsmoothed noisy data, although this problem disappears in ideal noise-free data if the sampling problems are handled properly. This scale determining the amount of *initial smoothing* in the (traditional first-stage) multi-scale representation of the image is referred to as *local scale*<sup>3</sup> (denoted  $t$ ).

### 3.1 Scale selection: Basic principle

Scale-space theory provides a methodology for handling such size or scale variations in image data; for a summary of main results, see e.g. (Witkin 1983; Koenderink 1984; Babaud et al 1986; Yuille and Poggio 1986; Lindeberg 1990, 1992; Koenderink and van Doorn 1990; Florack et al 1992). Given a two-dimensional continuous signal  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , the scale-space representation  $L : \mathbb{R}^2 \times \mathbb{R}_+ \rightarrow \mathbb{R}$  is defined as the solution to the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L = \frac{1}{2} (\partial_{xx} + \partial_{yy}) L \quad (14)$$

with initial condition  $L(\cdot; 0) = f(\cdot)$ , or equivalently, by convolution with the Gaussian kernel  $L(\cdot; t) = g(\cdot; t) * f(\cdot)$ , where

$$g(x, y; t) = \frac{1}{\sqrt{2\pi t}} e^{-(x^2+y^2)/(2t)}. \quad (15)$$

A well-known property of this representation is that the amplitude of spatial derivatives

$$L_{x^i y^j}(\cdot; t) = \partial_{x^i y^j} L(\cdot; t) = g_{x^i y^j}(\cdot; t) * f(\cdot) \quad (16)$$

in general *decrease with scale*. As an example of this consider, say, a sinusoidal input signal of some given frequency  $\omega_0$ ; for simplicity in one dimension,

$$f(x) = \sin \omega_0 x, \quad (17)$$

for which the solution to the (one-dimensional) diffusion equation is

$$L(x; t) = e^{-\omega_0^2 t/2} \sin \omega_0 x. \quad (18)$$

The amplitude  $L_{x^i, max}$  of any  $i$ th order smoothed derivative decreases exponentially with scale

$$L_{x^i, max}(t) = \omega_0^i e^{-\omega_0^2 t/2}. \quad (19)$$

An alternative formulation of the scale-space concept is in terms of *normalized* (dimensionless) *coordinates*,  $\xi = x/\sqrt{t}$ . One motivation for introducing such coordinates

---

<sup>3</sup>This terminology refers to local operations (derivatives).

is *scale invariance* (Florack et al 1992). In these coordinates the *normalized derivative operator* is

$$\partial_{\xi} = \sqrt{t} \partial_x. \quad (20)$$

For the sinusoidal signal the amplitude of a normalized derivative as function of scale is given by

$$L_{\xi^i, max}(t) = t^{i/2} \omega_0^i e^{-\omega_0^2 t/2}, \quad (21)$$

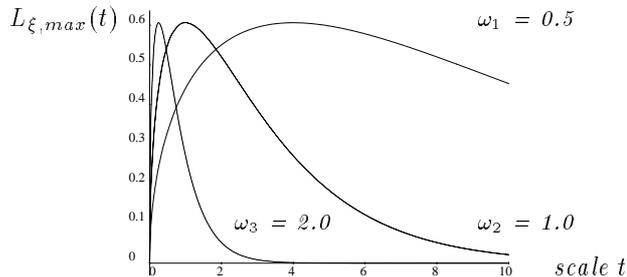
i.e., it first increases and then decreases. It assumes a *unique* maximum at  $t_{max, L_{\xi^i}} = i/\omega_0^2$ . Introducing  $\lambda_0 = 2\pi/\omega_0$  shows that scale value (measured in  $\sqrt{t}$ ) for which  $L_{\xi^i, max}(t)$  assumes its maximum is *proportional to the wavelength*,  $\lambda_0$ , of the signal:

$$\sqrt{t_{max, L_{\xi^i}}} = \frac{\sqrt{i}}{2\pi} \lambda_0. \quad (22)$$

Observe that the maximum value

$$L_{\xi^i, max}(t_{max, L_{\xi^i}}) = i^{i/2} e^{-i/2} \quad (23)$$

is *independent of the frequency* of the signal (see Figure 2). In other words, for these normalized derivatives sinusoidal signals are treated in a similar and scale invariant way independent of their frequency.



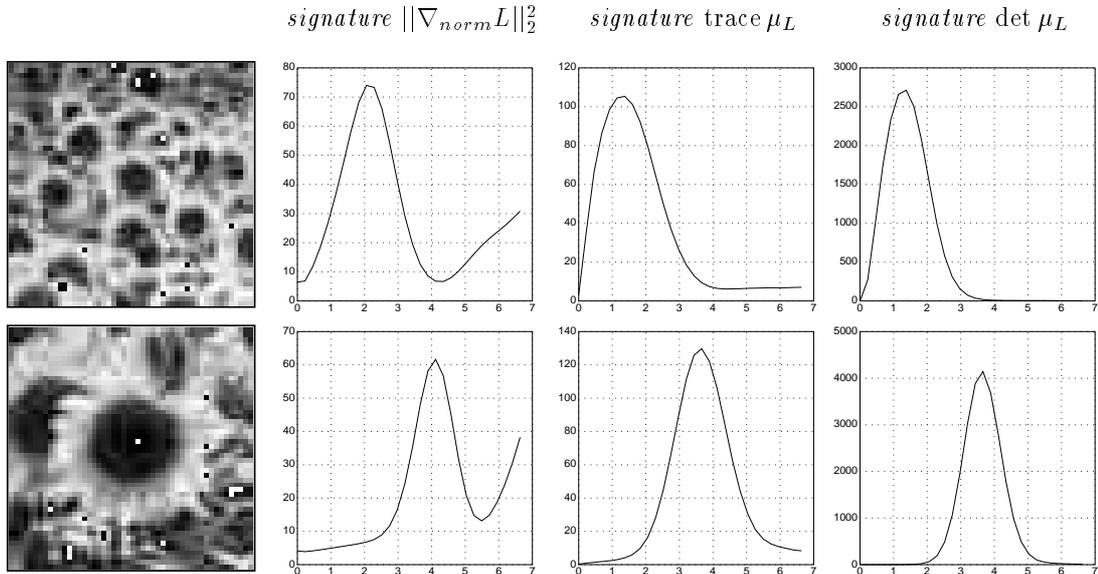
**Figure 2:** The amplitude of first order normalized derivatives as function of scale for sinusoidal input signals of different wavelengths ( $\omega_1 = 0.5$ ,  $\omega_2 = 1.0$  and  $\omega_3 = 2.0$ ).

Note the fundamental differences compared to a local Fourier transform; (i) the normalization factor, and (ii) this method allows for local estimates of the frequency content without any explicit setting of window size.

### 3.2 Proposed method for scale selection

As shown above, the scale at which the normalized derivative assumes its maximum in the case of a sinusoidal signal is proportional to the wavelength of the signal. We now propose to generalize this observation to more complex signals, leading to the following heuristic principle:

In the absence of other evidence, a scale level at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum can be treated as a characteristic dimension of a corresponding structure contained in the data.



**Figure 3:** Scale-space signatures of the pointwise and integrated normalized gradient magnitude ( $\|\nabla_{norm} L\|_2^2$  and  $\text{trace } \mu_L$  respectively), as well as the determinant of the second moment matrix ( $\det \mu_L$ ) for two details of a sunflower image; (left) grey-level image, (middle left) signature of  $\|\nabla_{norm} L\|_2^2$ , (middle right) signature of  $\text{trace } \mu_L$ , and (right) signature of  $\det \mu_L$ . Observe that the maxima in the row are assumed at finer scales than the maxima in the bottom row. (All entities are computed at the central point. The scaling of the horizontal axis is basically logarithmic, while the scaling of the vertical axis is linear.)

This principle is similar although not equivalent to the method for scale selection proposed by Lindeberg (1991), who selected interesting scales from maxima over scales of a normalized measure of the strength of a blob response in scale-space. It can be theoretically justified for a number of specific local brightness models, see Section 3.3 and (Lindeberg and Gårding 1993), but its general usefulness must be verified empirically.

Figure 3 illustrates the variation over scale of three simple measures formulated in terms of normalized spatial derivatives and computed at two different points. First, the scale variation of the normalized square of the *gradient magnitude*,  $\|\nabla_{norm} L\|_2^2$ , is shown. Second, the local average of the gradient magnitude computed using a Gaussian window function with the integration scale proportional to the local scale is shown; this entity is the *trace* of the windowed second moment matrix  $\mu_L(q)$ . Third, the *determinant* of  $\mu_L(q)$  is shown. These graphs are called the *scale-space signatures* of the entities considered.

Clearly, the maxima over scales in the top row of Figure 3 are obtained at finer scales than in the bottom row. An examination of the ratio between the scale levels where the graphs attain their maxima shows that this value is roughly equal to the ratio of the sizes of the sunflowers in the centers of the two images respectively, as predicted by the heuristic principle.

It should be pointed out that this principle for scale selection is not restricted to texture analysis; see (Lindeberg 1993) for further applications to junction detection

and edge detection.

### 3.3 Properties of the scale selection method

We will now describe some properties of the scale selection heuristic for slightly more complex signals. A more extensive treatment is given in the references cited above.

Consider first a sum of two *parallel* (two-dimensional) sine waves.

$$f_{par}(x, y) = \sin \omega_1 x + \sin \omega_2 x, \quad (24)$$

where  $\omega_1 \leq \omega_2$ . It is easy to show that for both  $\|\nabla_{norm} L\|_2^2$  and  $\text{trace } \mu_L$  there is a unique scale maximum when  $\omega_2/\omega_1$  is close to one, while there are two scale maxima for sufficiently large  $\omega_2/\omega_1$  ( $\omega_{bifurc} \approx 2.4$ ). A similar result holds for two *orthogonal* waves,

$$f_{orth}(x, y) = \sin \omega_1 x + \sin \omega_2 y. \quad (25)$$

If the latter signal is interpreted as the orthographic projection of an isotropic pattern with foreshortening  $\epsilon = \omega_1/\omega_2$ , then the interpretation is that the response changes from one to two peaks at slant  $\sigma_{bifurc} = \arccos(1/\omega_{bifurc}) \approx 65^\circ$ .

The determinant of the windowed second moment matrix,  $\det \mu_L$ , behaves somewhat differently; it is identically zero for  $f_{par}$ , while there is *always* a unique peak in  $f_{orth}$ .

More generally, for an *isotropic* pattern (with  $\tilde{Q} = 0$ , or equivalently,  $\lambda_1 = \lambda_2$ ) the scale maxima of  $\text{trace } \mu_L$  and  $\det \mu_L$  coincide. This is easily proved from  $\text{trace } \mu_L = \lambda_1 + \lambda_2 = 2\lambda_1$  and  $\det \mu_L = \lambda_1 \lambda_2 = \lambda_1^2$ , which gives  $\partial_t \det \mu_L = 0 \Leftrightarrow \partial_t \text{trace } \mu_L = 0$ .

For a *unidirectional* pattern (with  $\tilde{Q} = 1$ , or equivalently,  $\lambda_2 = 0$ )  $\det \mu_L$  is identically zero, while  $\text{trace } \mu_L$  is non-zero. Hence,  $\det \mu_L$  only responds when there are significant variations along *both* the coordinate directions, typically for blob-like signals.

The behaviour of the normalized derivatives can be understood also in the context of signals having a dense Fourier spectrum. For a signal  $f$  with a (fractal) power spectrum  $\Phi_f = \hat{f} \hat{f}^* = |\omega|^{-2\alpha}$  it follows from Plancherel's relation that

$$P_{norm}(\cdot; t) = t(E(L_x^2(\cdot; t)) + E(L_y^2(\cdot; t))) \sim t^{\alpha-1}. \quad (26)$$

This expression is independent of scale if and only if  $\alpha = 1$ . In other words, in the two-dimensional case the normalized derivative model is *neutral* with respect to power spectra of the form  $|\omega|^{-2}$ .

### 3.4 Uniqueness of the window function

We have suggested that the Gaussian is a natural choice of window function in (2). This choice could in principle be motivated by the fact that this kernel is rotationally symmetric with a nice scaling behaviour, which means that the invariance properties described in Section 2.1 are preserved. More importantly, however, it holds that *if and only if* the window function is a Gaussian, then the components of  $\mu_L$ ,  $\mu_{ij}$ , constitute *scale-space representations* of the components of  $(\nabla L)(\nabla L)^T$ ,  $L_{x_i} L_{x_j}$ , respectively. This is a direct consequence of the uniqueness of the Gaussian kernel

for scale-space representation given natural front-end postulates (e.g. the causality condition introduced by Koenderink (1984), or the scale invariance used by Florack et al (1992)). In the rotationally symmetric case, this formal definition of the multi-scale window second moment matrix  $\mu_L : \mathbb{R}^2 \times \mathbb{R}_+^2 \rightarrow \text{SPSD}(2)$  (obtained from the scale-space representation  $L$  of a signal  $f$ ) is therefore unique:

$$\mu_L(\cdot; t, s) = g(\cdot; s) * ((\nabla L)(\cdot; t) (\nabla L)(\cdot; t)^T). \quad (27)$$

Of course, separate smoothing of the components of a multi-dimensional entity is not guaranteed to give well-defined (coordinate independent) results. In (Lindeberg and Gårding 1993) it is, however, proved that (27) is a meaningful operation.

### 3.5 Scale selection for computation of $\mu_L$

Computation of the windowed second moment matrix  $\mu_L$  requires selection of both the local scale  $t$  and the integration scale  $s$ . In its most general form, the adaptive scheme we propose for setting these scales can be summarized as follows. Given any point in the image;

1. vary the two scale parameters, the local scale  $t$  and the integration scale  $s$ , according to some scheme;
2. accumulate the scale-space signature for some (normalized) differential entity;
3. detect some special property of the signature, e.g., the global maximum, or all local extrema, etc;
4. set the integration scale(s) used for computing  $\mu_L$  proportional to the scale(s) where the above property is assumed;
5. compute  $\mu_L$  at the fixed integration scale while varying the local scale between a minimum scale, e.g.  $t = 0$ , and the integration scale, and then select the most appropriate local scale(s) according to some criterion.

Our specific implementation of this general scheme is described below.

**Scale variation.** A completely general implementation of Step 1 would involve a full two-parameter scale variation. Here, a simpler but quite useful approach is used; the integration scale is set to a constant times the local scale,  $s = \gamma_1^2 t$  (typically  $\gamma_1 = 2$ ). In light of the scale selection heuristic, this scale invariant choice means that the size of the integration region is proportional to the characteristic length of the local smoothing kernel. For example, in the case of periodic patterns, this implies that the size of the integration region at each local scale is proportional to the wavelength for which the normalized first derivative at that scale would give a maximum response.

**Selecting integration scales.** Concerning Steps 2–3, we propose to set the integration scales from the scales, denoted  $s_{\det \mu_L}$ , where the normalized strength of  $\mu_L$ , represented by  $\det \mu_L$ , assumes a local or global maximum. This choice is motivated by the observation that for both simple periodic and blob-like patterns, the signature of  $\det \mu_L$  has a single peak reflecting the characteristic size (area) of the two-dimensional pattern, while for the pointwise and integrated gradient magnitude the response changes from one to two peaks with increasing (linear) distortion.

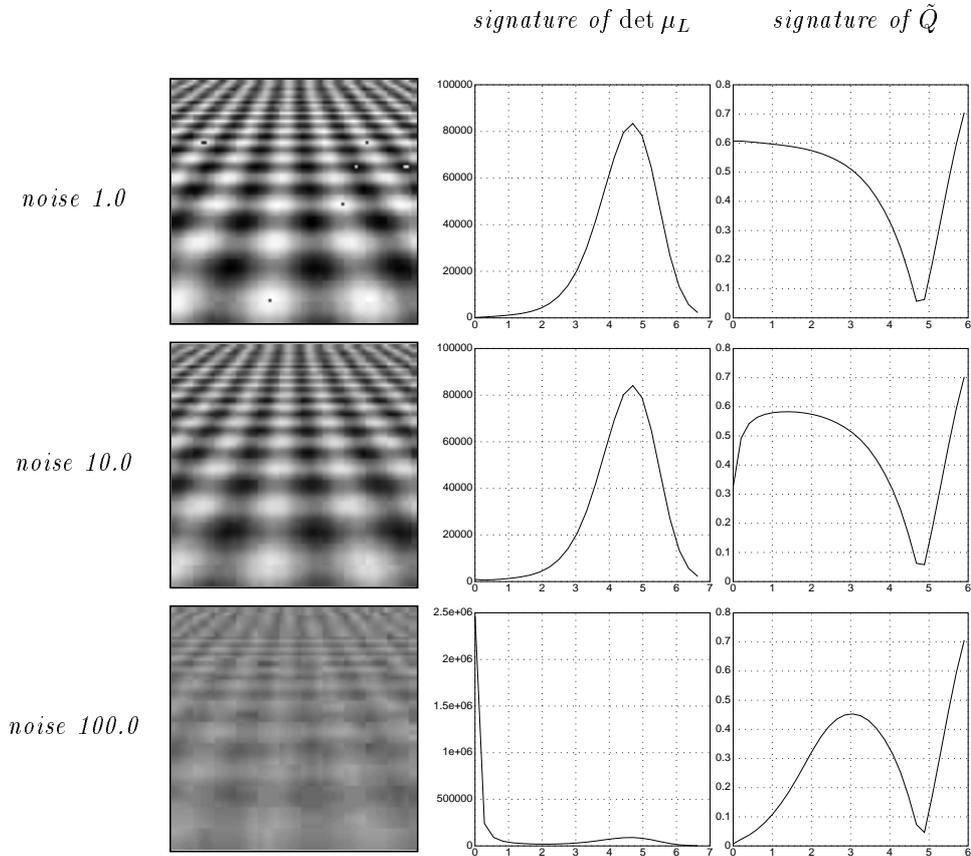
Once  $s_{\det \mu_L}$  has been determined, it is advantageous to compute  $\mu_L$  at a slightly larger integration scale  $s = \gamma_2^2 s_{\det \mu_L} = \gamma_1^2 \gamma_2^2 t_{\det \mu_L}$  (typically  $\gamma_2 = 2$ ), in order to obtain a more stable descriptor. More formally, using  $\gamma_2 > 1$  can be motivated by the analysis in (Lindeberg and Gårding 1993), which shows that the estimates of the directional information in  $\mu_L$  are more sensitive to small window sizes than are the magnitude estimates.

**Selecting local scales.** The second stage selection of local scale in Steps 5–6 aims at reducing the shape distortions due to smoothing. We propose to set the *local scales* to the scales, denoted  $t_Q$ , where the *normalized anisotropy*,  $\tilde{Q}$ , assumes a local maximum. This is motivated by the fact that in the absence of noise and interfering finer scale structures, the main effect of the first stage scale-space smoothing is to *decrease* the anisotropy. For example, the aspect ratio of an elliptical Gaussian blob  $f(x, y) = g(x; l_1^2) g(y; l_2^2)$  varies as  $(l_2^2 + t)/(l_1^2 + t)$ , and clearly approaches one as  $t$  is increased. On the other hand, suppressing isotropic noise and interfering finer scale structures *increases* the anisotropy. Selecting the maximum point gives a natural trade-off between these two effects.

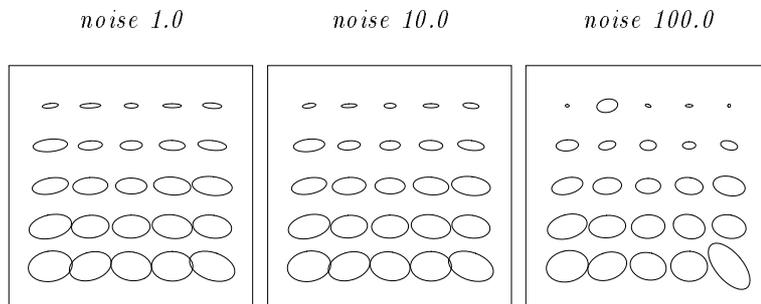
**Experiments.** Figure 3 illustrates these effects for a synthetic image with different amounts of additive white Gaussian noise. Note that the scale-space signature of  $\det \mu_L$  has a unique maximum when the noise level,  $\nu$ , is small, and two maxima when  $\nu$  is increased. Table 1 gives numerical values obtained by using the proposed method for scale selection. Notice the stability of  $s_{\det \mu_L}$  with respect to noise. The selected local scale  $t_Q$  increases with the noise level  $\nu$ , while  $\tilde{Q}$  decreases at  $t = 0$ .

noise level	$s_{\det \mu_L}$	$t_Q$	$\tilde{Q}(t_Q)$	$\tilde{Q}(t = 0)$	$\Delta\phi_n(t_Q)$	$\Delta\phi_n(t = 0)$
1.0	34.9	0.0	0.602	(0.602)	0.2°	(0.2°)
10.0	34.4	2.0	0.579	(0.329)	1.1°	(15.3°)
31.6	34.1	4.2	0.510	(0.033)	4.7°	(45.3°)
100.0	31.4	8.5	0.456	(0.006)	7.8°	(53.7°)

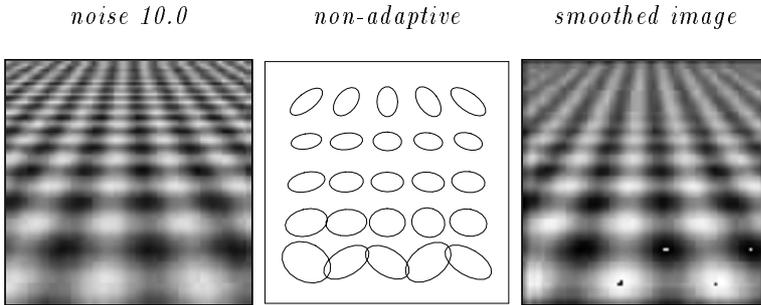
**Table 1:** Numerical values of some characteristic entities in the experiments in (the center of) Figure 4 using different amounts of additive Gaussian noise and automatic scale selection. Note the stability of the selected integration scale (proportional to  $s_{\det \mu_L}$ ) with respect to variations in the noise level  $\nu$ , and that the selected local scale  $t_Q$  increases with  $\nu$ . Observe also the increasing difference between the estimates of the normalized anisotropy  $\tilde{Q}$  computed at the selected local scale, and at zero local scale (true value 0.600). The last two columns show the error in surface orientation  $\Delta\phi_n$  computed by monocular shape-from-texture under a specific assumption about the surface texture (weak isotropy).



**Figure 4:** Scale-space signatures of  $\det \mu_L$  and  $\tilde{Q}$  (accumulated at the central point) for a synthetic texture with added (white Gaussian) noise of standard deviation  $\nu = 1.0$  (top row), 10.0 (middle row), and 100.0 (bottom row). The range of grey-levels is  $[0..255]$ . The columns show; (left) grey-level image with noise, (middle) signature of  $\det \mu_L$ , and (right) signature of  $\tilde{Q}$ .



**Figure 5:** Ellipses representing  $\mu_L$  computed at different spatial points using *automatic scale selection* of the local scale and the integration scale — note the stability with respect to variations of the noise level.



**Figure 6:** Typical example of the result of using *non-adaptive* selection of the (here constant) local and integration scales — geometrically useful shape descriptors are obtained only in a small part of the image.

In Section 5.3 it is shown that under a certain assumption about the surface texture (weak isotropy), the estimate of surface orientation is directly related to the normalized anisotropy  $\tilde{Q}$ , and to the eigenvector of  $\mu_L$  corresponding to the maximum eigenvalue. Table 1 illustrates the accuracy in estimates of  $\tilde{Q}$  and surface orientation computed in this way. The error in surface orientation is measured by the three-dimensional angle  $\Delta\phi_n$  between the estimated and true surface normal.

Figure 5 illustrates these results graphically, by ellipses representing the second moment matrices, with the size rescaled to be proportional to  $s_{\det \mu_L}$ . As a comparison, Figure 6 displays a typical result of using non-adaptive (globally constant) scale selection. Here, useful shape descriptors are only obtained in a small part; the window size is too small in the lower part, while the first stage smoothing leads to severe shape distortions in the upper part.

### 3.6 The ellipse representation revisited

The ellipse given by (4) graphically represents the local statistics of the first-order directional derivatives computed at the local scale  $t$  and the integration scale  $s$ . In particular, the area  $A = 1/\sqrt{\det \mu_L}$  of the ellipse reflects the average magnitude of these derivatives, and is unrelated to the characteristic dimension(s) of the image structures. This can be seen e.g. by noting that scaling of the image brightness by some factor  $k$  scales  $A$  by  $1/k^2$ , whereas the shape of the ellipse remains unchanged. Hence, ellipses computed in a dim region of the image on average tend to be larger than those computed in areas of higher contrast. For this reason, the information about local image structure contained in the absolute magnitude of the components of  $\mu_L$  is somewhat unreliable, and in our present implementation we have chosen not to use it at all.

On the other hand, reliable information about the characteristic size of image structure is available from the scale selection procedure. We therefore normalize  $\mu_L$  by scaling its components to make the area of the ellipse proportional<sup>4</sup> to the scale

<sup>4</sup>The scale factor is selected such that for a circular binary blob the ellipse area is equal to the area of the blob. This only affects how ellipses are displayed; in the computations of various shape cues from  $\mu_L$  the scale factor always cancels out.

at which the maximum of  $\det \mu_L$  is assumed.

Finally, note that the resulting normalized second moment descriptor is independent of the absolute strength of the response, measured by  $\det \mu_L$  or  $\text{trace} \mu_L$ . The same holds for  $\det \mathcal{H}_{norm} L$  and  $\text{trace} \mathcal{H}_{norm} L$ . This makes it possible to use these measures for further independent classification of the local brightness pattern in terms of, e.g., contrast.

## 4 Spatial selection and blob detection

The previous sections treated the problem of selecting appropriate scales for local smoothing and regional integration at a given image point. In this section the complementary problem of selecting *where* in the image to apply the multi-scale analysis will be addressed. This problem will be referred to as *spatial selection*, to emphasize the analogy with scale selection.

Spatial selection could in principle be avoided by computing a texture descriptor at every image point, but this is typically not an acceptable solution; it can lead to unnecessarily poor estimates since many image points often contain little or no useful image structure.<sup>5</sup> In particular, many natural textures seem to consist of fairly similar texture elements randomly scattered on the surface. This is quite unlike the idealized case of a perfectly periodic texture, in which all image points provide more or the less the same information as long as the integration scale is proportional to the wavelength of the projected texture.

It will now be shown that the proposed scale selection method can be successfully extended to guide a spatial selection process as well. The resulting simultaneous selection of scale and spatial position can be interpreted as a form of *multi-scale blob detector*, where each detected blob is represented by its position, its detection scale, and a second moment matrix. This multi-scale blob detector has obvious limitations compared to more general approaches, e.g. (Blostein and Ahuja 1989; Lindeberg and Eklundh 1992), since it only represents the shape of each blob by a second moment matrix. However, we propose that it is well suited as a pre-processing step for the shape estimation processes described in Sections 5 and 6, since it produces precisely the information needed for estimating local linear distortion and size changes.

### 4.1 Spatial selection: Basic principle

In Section 3.5 we proposed to select scales at a given image point by the local maxima over scale of some (possibly non-linear) combination of normalized spatial derivatives. This principle can be generalized to achieve spatial selection as well, by selecting points  $(x, y)^T$  and scales  $t$  that are local maxima with respect to scale *and* position of such an entity. Such points are denoted *normalized scale-space maxima* of the differential entity considered.

The most straightforward implementation of this general principle is to use the same normalized entity for spatial selection as was used in the selection of integration scale, i.e.,  $\det \mu_L$  (see Section 3.5). This method has the advantage that spatial

---

<sup>5</sup>When implementing the algorithm on a serial computer there are obviously efficiency considerations as well.

selection and scale selection are performed simultaneously. Alternatively, the spatial selection can be performed independently of the scale selection. In particular, it may be desirable to use an operator based on second order derivatives (even operators), since such an operator typically gives rise to spatial maxima at the centers of high contrast blobs that stand out from the surrounding.

Previous methods for blob detection have often been based on the Laplacian of the Gaussian,  $\nabla^2 g$ ; see e.g (Marr 1982; Blostein and Ahuja 1989a, 1989b; Voorhees and Poggio 1987). It is common for methods utilizing  $\nabla^2 g$  or similar operators to be combined with some thresholding operation in order to suppress false alarms, and also to contain a more or less complex spatial post-processing step, in which blobs may, e.g., be split or merged according to some geometric criterion. In contrast, the scheme we propose contains neither thresholding nor spatial post-processing.

For the purpose of spatial selection, we have investigated the use of three different non-linear combinations of normalized derivatives, all of them well-defined in the sense that they do not depend on the choice of coordinate system:

- The determinant of the second moment matrix,  $\det \mu_L$ , i.e., the same property as was favoured for scale selection previously.
- The squared<sup>6</sup> Laplacian  $(L_{\xi\xi} + L_{\eta\eta})^2$ , i.e., the squared trace of the normalized Hessian,  $\text{trace}^2 \mathcal{H}_{norm} L$ .
- The determinant of the normalized Hessian matrix,  $\det \mathcal{H}_{norm} L = L_{\xi\xi} L_{\eta\eta} - L_{\xi\eta}^2$ . This is the normalized Gaussian curvature of the brightness surface multiplied by a factor that depends on the magnitude of the gradient.

An analysis concerning the scales at which these entities assume local maxima over scales for a periodic and a blob-like pattern respectively is given in (Lindeberg and Gårding 1993); some results are summarized in Table 2. Note that the scales at which the maxima are assumed are related by constant factors.

Model signal	$t_{\text{trace} \mu_L}$	$t_{\text{det} \mu_L}$	$t_{\text{trace} \mathcal{H}_{norm} L}$	$t_{\text{det} \mathcal{H}_{norm} L}$
Periodic: $\sin \omega_1 x + \sin \omega_2 y$	$1/\omega_0^2$	$2/(\omega_1^2 + \omega_2^2)$	$2/\omega_0^2$	$4/(\omega_1^2 + \omega_2^2)$
Blob: $g(x; t_1) g(y; t_2)$	$t_0/\sqrt{1 + 2\gamma_1^2}$	$\sqrt{t_1 t_2}/\sqrt{1 + 2\gamma_1^2}$	$t_0$	$\sqrt{t_1 t_2}$

**Table 2:** Analytical expressions for the scale levels where the local maxima over scales are attained for a periodic model signal and a blob-like model signal. For the entities based on the trace of  $\mu_L$  and  $\mathcal{H}_{norm} L$  respectively, only the results from the isotropic cases ( $\omega_1 = \omega_2 = \omega_0$ , and  $t_1 = t_2 = t_0$ ) are shown. For the periodic signal the trace based entities have two extrema when the foreshortening is sufficiently large, while the maximum is unique for determinant based entities. For the blob signal, the maximum is unique in all four cases.

In practice, each of these entities is computed at an integration scale  $s = \gamma_1^2 t$  proportional to the local scale  $t$ . In the first case, the integration is applied to  $\mu_L$

<sup>6</sup>The squaring is performed only in order to obtain uniform treatment of bright and dark blobs. The same effect could, of course, also be achieved by considering both normalized scale-space maxima and normalized scale-space minima of the ordinary Laplacian operator (although the effect of the second stage smoothing then would become somewhat different).

before the determinant is taken, since  $\det \mu_L$  is identically zero when considered pointwise. In contrast, the pointwise representations of the other two operators are not singular, so in these cases the integration step could in principle be omitted (i.e.,  $\gamma_1 = 0$ ). Nevertheless, we have sometimes found it advantageous to apply the same type of integration to these operators as well; the main effect of this integration step is to suppress a large number of less significant responses, and hence to reduce the computational load.

## 4.2 Experimental results

The properties of the spatial selection process will now be illustrated using two synthetic test images. Additional experiments, using natural images, are given in Section 5.

The result of the first experiment is shown in Figure 7. The image to the left contains dark elliptical blobs with varying sizes and aspect ratios on a brighter background, and additive Gaussian noise with a standard deviation equal to 20% of the brightness difference between the blobs and the background. The blob positions detected by each of the three operators in this image are shown to the right.<sup>7</sup> Since no shape information is computed at this stage, the detected blobs are displayed as circles, with the area of each circle proportional to the detection scale.

The performance of all three operators is somewhat similar, but it is clear that they differ in the number of spurious maxima they generate, as well as in their tendency to generate multiple spatial maxima for elongated blobs. Clearly,  $\det \mu_L$  generates most maxima, and  $(\text{trace } \mathcal{H}_{norm} L)^2$ , i.e., the squared normalized Laplacian, generates the fewest. Subsequent experiments on spatial selection will therefore be based on the latter operator, but it should not be ruled out that the other two operators can be advantageous in some situations.

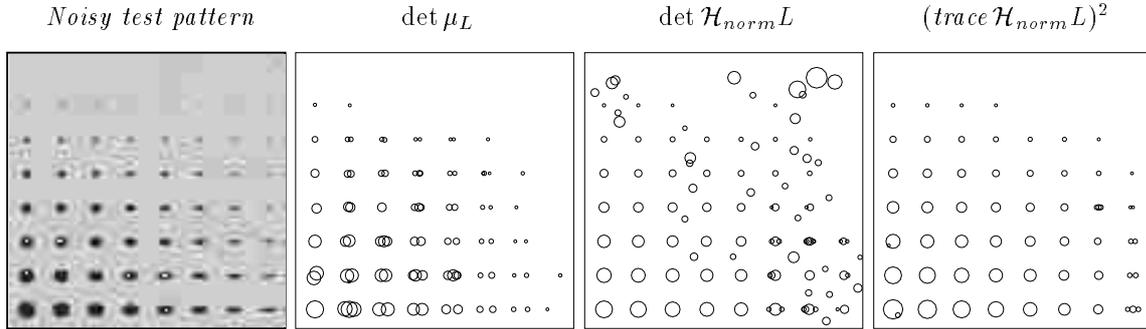
Figure 8 shows the final blobs found by the method, using the squared Laplacian for spatial selection and  $\det \mu_L$  for computation of blob size and shape as explained previously. In the scale selection step, the integration scale parameter was coupled to the local scale parameter by  $s = \gamma_1^2 t$  with  $\gamma_1 = \sqrt{2}$ . Then, when computing the second moment matrices, the integration scale was set to  $s = \gamma_2^2 s_{\det \mu_L}$  with  $\gamma_2 = \sqrt{2}$ , where  $s_{\det \mu_L}$  denotes the integration scale for which the maximum in  $\det \mu_L$  was assumed. Here, only the global maxima with respect to scale have been retained.

The last example of this section demonstrates the importance of adapting the local scale in the computation of  $\mu_L$ . Figure 9 shows the blobs detected in an image at the local scale that maximizes  $\det \mu_L$ , as well as the final blobs obtained by adapting the local scale to maximize anisotropy.

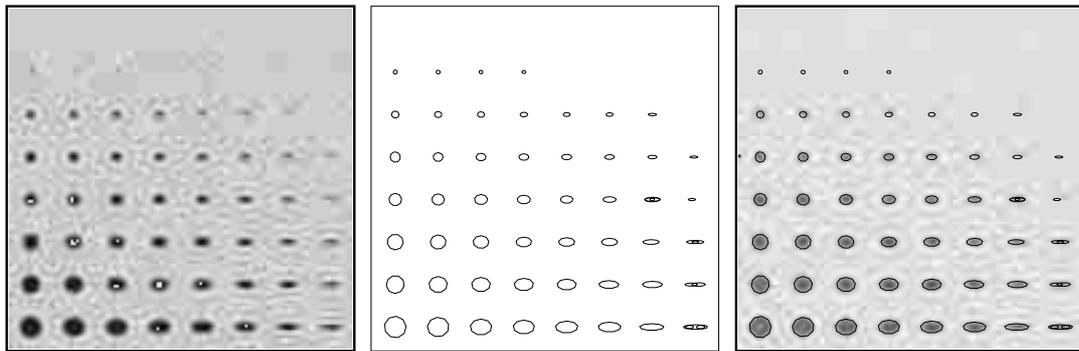
Finally, it is worth pointing out that although the spatial selection method was motivated from considerations of non-periodic textures, there is no reason why it cannot be applied to periodic textures as well. The fundamental limitation of the method is that the texture cannot be too anisotropic, because then the initial detection step based on the Laplacian is likely to fail. However, a blob-based description of such textures does not seem very meaningful anyway.

---

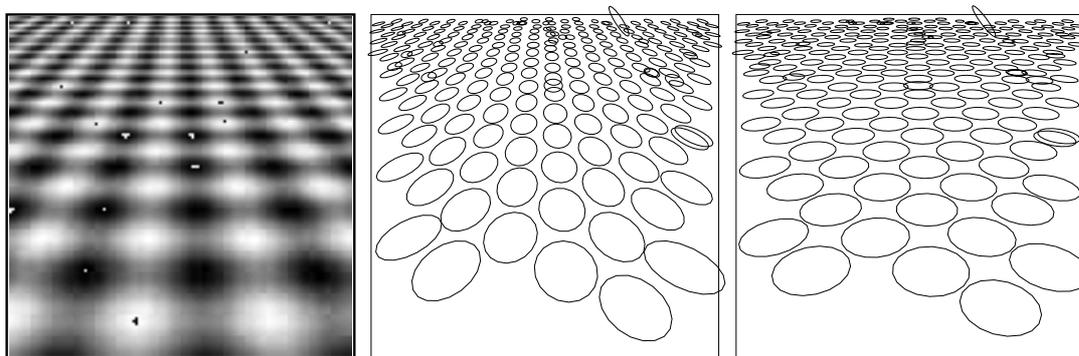
<sup>7</sup>The actual implementation here was based on eight scale levels, ranging from  $t = 1$  to  $t = 128$  and distributed in logarithmic steps. The integration scale was  $s = 2t$ , and the image size was  $512 \times 512$ .



**Figure 7:** (a) Synthetic image with dark elliptical blobs with varying sizes and aspect ratios on a brighter background, and additive Gaussian noise with a standard deviation equal to 20% of the brightness difference between the blobs and the background. (b)–(d) Blobs detected using integration scale  $s = \gamma_1^2 t$  with  $\gamma_1 = \sqrt{2}$ . From left to right, the operator used was  $\det \mu_L$ ,  $\det \mathcal{H}_{norm} L$ , and  $(\text{trace } \mathcal{H}_{norm} L)^2$ . The size of each circle indicates the scale at which the maximum was assumed.



**Figure 8:** Multi-scale blob detection using normalized scale-space extrema of the square of the Laplacian of the Gaussian. (Left) Original image. (Middle) Detected ellipses. (Right) Ellipses representing the second moment matrix superimposed onto a bright copy of the original grey-level image.



**Figure 9:** Multi-scale blob detection using normalized scale-space extrema of the square of the Laplacian of the Gaussian. (Left) Original image. (Middle) Detected ellipses before adaption of local scale. (Right) Detected ellipses after adaption of local scale.

## 5 Shape from texture

This section shows how the proposed multi-scale texture descriptor can be used to estimate the shape or orientation of three-dimensional surfaces in the scene from perspective distortion of surface texture observed in a monocular image.

### 5.1 Background

The image of a slanted textured surface contains several more or less independent cues that can be used to estimate the shape and orientation of the surface. Pioneering work was done by Gibson (1950), who studied so-called texture gradients, i.e., systematic variations in the image texture due to perspective effects. One example is the familiar “perspective effect” which makes the image of a near surface patch smaller than that of a far patch. Several algorithms for estimation of surface orientation from texture gradients have later been proposed, e.g. (Aloimonos 1988; Blostein and Ahuja 1989; Kanatani and Chou 1989; Blake and Marinos 1990a).

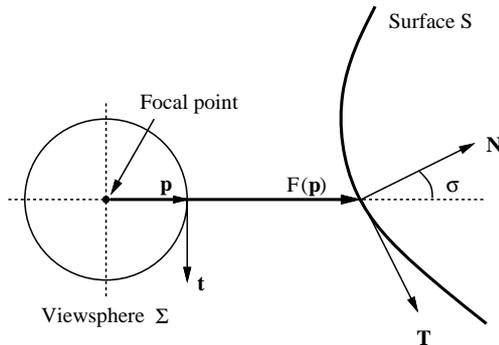
Witkin (1981) pointed out that the foreshortening effect, i.e., the systematic compression of a slanted pattern in the direction of slant, can also be a cue to surface orientation. For example, the image of a slanted circle is an ellipse, and the degree and orientation of the elongation of the ellipse indicates the magnitude and direction of slant. Whereas texture gradients are primarily due to perspective effects, the foreshortening effect can also be observed in orthographic projection of a planar pattern. Various extensions of Witkin’s method have later been described, e.g. (Davis et al 1983; Kanatani 1984; Blake and Marinos 1990; Gårding 1993). Related methods include (Pentland 1986; Brown and Shvaytser 1990).

### 5.2 Review of image geometry

In order to understand how a local texture description can be interpreted in terms of three-dimensional surface shape, it is necessary to take a closer look at the surface and viewing geometry.

Consider a smooth surface  $S$  viewed in perspective projection. The local perspective distortion of the projected surface pattern results from two factors; firstly, the distance and orientation of the surface with respect to the line of sight, and secondly, the angle between the line of sight and the image surface. The latter factor is often referred to as the “position effect”. Since it only depends on the internal camera geometry it can be eliminated by reprojection of the image from the focal point. Hence, for analytical clarity we represent the image by a unit viewsphere  $\Sigma$ , and let it be understood that in practical computations with a planar image the coordinates on  $\Sigma$  are obtained by a local coordinate transformation.

Fortunately, it can be shown that in order to estimate local surface shape from texture, it suffices to consider the first-order (linear) terms of the perspective projection at each image point. To give a more precise formulation of this statement, it is necessary to introduce a few definitions. Following (Gårding 1992), and using standard notation from differential geometry (see e.g. (O’Neill 1966)), consider a spherical camera mapping a smooth surface  $S$  onto a unit viewsphere  $\Sigma$  (see Figure 10). At any point  $p$  on  $\Sigma$  let  $(\bar{p}, \bar{t}, \bar{b})$  be a local orthonormal coordinate system defined such



**Figure 10:** Local surface geometry and imaging model. The tangent planes to the viewsphere  $\Sigma$  at  $p$  and to the surface  $S$  at  $F(p)$  are seen edge-on but are indicated by the tangent vectors  $\bar{t}$  and  $\bar{T}$ . The tangent vectors  $\bar{b}$  and  $\bar{B}$  are not shown but are perpendicular to the plane of the drawing, into the drawing. (Adapted from (Gårding 1992)).

that the  $\bar{p}$  direction is parallel to the view direction,  $\bar{t}$  is parallel to the direction of the gradient of the distance from the focal point, and  $\bar{b} = \bar{p} \times \bar{t}$ . Denote by  $F : \Sigma \rightarrow S$  the perspective backprojection from  $\Sigma$  to  $S$ , and by  $F_{*p}$  the derivative of this mapping at any point  $p$  on  $\Sigma$ . The mapping  $F_{*p}$ , which constitutes a linear approximation of  $F$  at  $p$ , maps point in the tangent plane of  $\Sigma$  at  $p$ , denoted  $T_p(\Sigma)$ , to points in the tangent plane of  $S$  at  $F(p)$ , denoted  $T_{F(p)}(S)$ . In  $T_{F(p)}(S)$ , let  $\bar{T}$  and  $\bar{B}$  be the normalized images of  $\bar{t}$  and  $\bar{b}$  respectively. In the bases  $(\bar{t}, \bar{b})$  and  $(\bar{T}, \bar{B})$  the expression for  $F_{*p} : T_p(\Sigma) \rightarrow T_{F(p)}(S)$  is

$$F_{*p} = \begin{pmatrix} r / \cos \sigma & 0 \\ 0 & r \end{pmatrix} = \begin{pmatrix} 1/m & 0 \\ 0 & 1/M \end{pmatrix}, \quad (28)$$

where  $r = \|F(p)\|$  is the distance along the visual ray from the center of projection to the surface (measured in units of the focal length) and  $\sigma$  is the slant of the surface. Two *characteristic* (dimensionless) *ratios* ( $m, M$ ) have been introduced to simplify later expressions and because of their geometric significance. These entities are the inverse eigenvalues of  $F_{*p}$ , and they basically describe how a unit circle in  $T_{F(p)}(S)$  is transformed when mapped to  $T_p(\Sigma)$  by  $F_{*p}^{-1}$ ; it becomes an ellipse with  $m$  as minor axis (parallel to the  $t$  direction) and  $M$  as major axis (parallel to the  $b$  direction).

From  $F_{*p}$  several useful relations between local perspective distortion and surface shape can be derived. Firstly, surface orientation is directly related to  $(m, M)$  and the corresponding eigenvectors  $(\bar{t}, \bar{b})$ . The *tilt* direction, defined as the direction of the gradient of the distance from  $\Sigma$  to the surface, is parallel to the eigenvector  $\bar{t}$  corresponding to the smaller inverse eigenvalue  $m$ . *Foreshortening* is defined as the ratio  $m/M$ , and is directly related to surface slant  $\sigma$  by the relation  $\cos \sigma = m/M$ . Together, tilt  $\bar{t}$  and slant  $\sigma$  determine the surface orientation (up to the sign of tilt; both  $\bar{t}$  and  $-\bar{t}$  are eigenvectors corresponding to the eigenvalue  $1/m$ ). Secondly, “texture gradients” can be computed from the spatial rate of change of various measures derived from the eigenvalues/eigenvectors of  $F_{*p}$ . For example, the local area ratio between the image and the surface is  $1/\det F_{*p} = mM$ , and the normalized *area gradient* which contains information about surface shape and

orientation is thus  $\nabla(mM)/(mM)$ . In Section 5.3 we will return to these relations, and show how they can be exploited in practice.

Normally, the brightness data are available in a planar image  $\Pi$ , rather than in the viewsphere  $\Sigma$ . This is of little consequence, however, because the mapping  $G : \Pi \rightarrow \Sigma$  from a point  $q$  on the planar image to the corresponding point  $p$  on the viewsphere can be pre-computed as long as the internal camera geometry is known.<sup>8</sup> The mapping  $F$  is then replaced by the composed mapping  $A = F \circ G$ , and the derivative map  $F_{*p}$  is replaced by the composed derivative map  $A_{*q} = F_{*p} G_{*q}$ , where  $p = G(q)$  and  $q \in \Pi$ . Hence, because  $G_{*q}$  is known,  $F_{*p}$  can always be computed from  $A_{*q}$ .

A more detailed discussion of the shape cues that can be derived from the components of  $F_{*p}$  and its derivatives can be found in (Gårding 1992).

### 5.3 Deriving shape cues from the second moment descriptor

In order to use a texture description derived from a monocular image to infer properties of the surface geometry, it is necessary to introduce some assumptions about the surface texture. These assumptions can have many different forms. In this section two useful examples are considered; firstly, *isotropy*, which allows estimation of “shape from foreshortening”, and secondly, *constant size*, which allows estimation of “shape from the area gradient”.

The derivations will be made in terms of the texture descriptor  $\mu_L(q)$ , which describes the windowed brightness structure in the image plane, and the corresponding descriptor  $\mu_\Sigma(p)$  defined in the tangent plane  $T_p(\Sigma)$  to the unit viewsphere  $\Sigma$  at the point  $p = G(q)$ . More precisely,  $\mu_\Sigma(p)$  describes the structure of the intensities transformed from the image to  $T_p(\Sigma)$  by the linearized mapping  $G_{*q}$ , and weighted by the transformed window function  $w'(p) = w(G_{*q}^{-1}p)$ . By (7) we have

$$\mu_L(q) = G_{*q}^T \mu_\Sigma(p) G_{*q}, \quad (29)$$

where  $G_{*q} : T_q(\Pi) \rightarrow T_p(\Sigma)$  is the derivative map between (the tangent plane to) the planar image  $\Pi$  at  $q$  and the tangent plane to the viewsphere at  $p$ . Hence, the practical procedure is to first estimate  $\mu_L(q)$  in the image plane and then to compute  $\mu_\Sigma(p)$  by inverting (29).

Analogously,  $\mu_S(F(p))$  describes the structure of the intensities transformed from  $T_p(\Sigma)$  by the linearized mapping  $F_{*p}$ , and weighted by the window function transformed accordingly. Assuming that the image brightness is directly proportional to the surface reflectance, it holds that  $\mu_S(F(p))$  describes the structure of the linearized and windowed surface reflectance at the point  $F(p)$ . Again, by (7) we have

$$\mu_\Sigma(p) = F_{*p}^T \mu_S(F(p)) F_{*p}. \quad (30)$$

The general procedure, then, for estimation of shape from texture is to combine estimates of  $\mu_L(q)$  (and possibly its derivatives) in the image plane with assumptions about the structure of the surface reflectance pattern  $\mu_S(F(p))$  in order to infer the structure of  $F_{*p}$  from (29) and (30).

To simplify the notation, the arguments to  $\mu_L$ ,  $\mu_\Sigma$  and  $\mu_S$  will be dropped in the remainder of this section.

---

<sup>8</sup>The mapping  $G$  is often referred to as the *gaze transformation*.

### 5.3.1 Shape from foreshortening

A simple but often fruitful assumption is that  $\mu_S$  is proportional to the unit matrix, i.e., that

$$\mu_S = cI \quad (31)$$

for some (unknown) constant  $c > 0$ . Such a distribution (for which  $\tilde{Q}_S = 0$ ) is called *weakly isotropic*. As described above, this essentially means that there is no single preferred direction in the surface texture, i.e., that the surface texture is not systematically elongated. Under this condition and assuming that  $F_{*p}$  is non-degenerate, (30) can be rewritten as

$$\mu_\Sigma = c F_{*p}^T F_{*p}. \quad (32)$$

Hence the eigenvectors of  $\mu_\Sigma$  and  $F_{*p}$  are the same, and the eigenvalues of  $F_{*p}$  are proportional to the square roots of the eigenvalues  $(\lambda_1, \lambda_2)$  of  $\mu_\Sigma$ ;

$$m \sim 1/\sqrt{\lambda_1} \sim 1/\sqrt{1 + \tilde{Q}}, \quad M \sim 1/\sqrt{\lambda_2} \sim 1/\sqrt{1 - \tilde{Q}}. \quad (33)$$

As shown in Sec. 5.2, the tilt direction,  $\bar{t}$ , is (plus/minus) the eigenvector,  $\bar{e}_1$ , corresponding to the maximum eigenvalue,  $\lambda_1$ , and the slant is given by

$$\cos \sigma = \frac{m}{M} = \sqrt{\frac{1 - \tilde{Q}}{1 + \tilde{Q}}} \quad (34)$$

Hence, if the assumption of weak isotropy can be justified, an easily computed estimate of local surface orientation is directly available. Unfortunately, many natural textures violate this assumption, and it is therefore often necessary to exploit alternative assumptions.

### 5.3.2 Shape from the area gradient

Assuming that the local “size” of the surface texture does not vary systematically, it is obvious that the gradient of size of the projected texture is an important cue to surface shape and orientation.

Consider a point  $p$  in  $T_p(\Sigma)$ , and let  $F_{*p}$  be the local linear part of the perspective backprojection. The area ratio is then equal to  $\det F_{*p}^{-1} = mM$ , i.e.,

$$A_\Sigma = mM A_S,$$

where  $A_\Sigma$  is the area of a small surface patch on the viewsphere  $\Sigma$ , and  $A_S$  is the area of the corresponding patch in the surface  $S$ . Hence, assuming that  $A_S = c$  where  $c$  is some unknown constant, we can define the *normalized area gradient*

$$\frac{\nabla A_\Sigma}{A_\Sigma} = \frac{\nabla(mM)}{mM}. \quad (35)$$

Note that the unknown scale constant  $c$  has been eliminated. This means that no assumptions about the absolute scale of the surface texture are necessary. Moreover, no assumptions are made about the elongation of the surface texture (given by the ratio of the eigenvalues of  $\mu_S$ ).

The area  $A_\Sigma$  can be computed from the corresponding area  $A_L$  in the planar image  $\Pi$  using  $A_\Sigma = (\det G_*)A_L$ , where  $G_*$  is the gaze transformation discussed in Section 5.2. In our current implementation  $A_L$  is estimated from the scale at which  $\det \mu_L$  assumes its maximum, as described in Section 3.6. A more detailed description of how to estimate the normalized area gradient from  $A_L$  is given in Appendix A.2.

It has not yet been mentioned how the normalized area gradient should be interpreted in terms of surface shape and geometry. It turns out that its information content is considerably more complex than that of foreshortening; in (Gårding 1992) it is shown that

$$\frac{\nabla(mM)}{mM} = -\tan \sigma \begin{pmatrix} 3 + r\kappa_t / \cos \sigma \\ r\tau \end{pmatrix}, \quad (36)$$

where  $r$  is the distance from the viewer,  $\sigma$  is the slant of the surface,  $\kappa_t$  is the normal curvature of the surface in the tilt direction, and  $\tau$  is the geodesic torsion, or “twist”, of the surface in the tilt direction.

Hence, the normalized area gradient can either be used to recover information about the surface curvature (scaled by distance) if the surface orientation is known, or to recover the surface orientation if the curvature is known or (assumed to be) small. In the latter case there is no ambiguity in the sign of the tilt direction, unlike the case of foreshortening.

#### 5.4 Estimating surface shape and orientation: Basic scheme

Our method for computing monocular shape-from-texture cues from image data can be summarized as follows:

1. Compute local texture descriptors  $\mu_L$  as described in Section 3.5. This can either be done at selected spatial positions corresponding to normalized scale-space extrema as described in Section 4, or at a (uniform) grid of points generated by some default principle.
2. Determine a set of points where estimates of surface orientation are to be computed. This set of points can be the same as that used for computing the texture descriptors, or it can be a smaller set of points, e.g. a uniform grid. Associate with each point a (Gaussian) window that specifies the weighting of the texture descriptors in the neighborhood of the point. The scale of this window function will be referred to as the *texel grouping scale*.
3. Estimate surface orientation:
  - (a) Apply the assumption of *weak isotropy* as described in Section 5.3 to compute foreshortening. This leads to a direct estimate of surface orientation up to the sign of tilt.
  - (b) Apply the assumption of *constant area* as described in Section 5.3 to compute the normalized area gradient. This permits a unique estimate of surface orientation under the additional assumption that the local curvature of the surface can be neglected in (36).
  - (c) Optionally, compute other texture gradients as well, e.g. the foreshortening gradient, and use them to estimate surface shape and/or orientation.

## 5.5 The texel grouping scale

In order to compute an estimate of surface orientation at a specified point, the local texture descriptors in the neighbourhood of the point must somehow be combined. As was described in previous sections, the second moment descriptor computed by spatial and scale selection can be informally thought of as a single “texture element”. In the case of a perfectly regular surface texture, the shape of this texture element can be relied upon to provide information about local perspective distortion. Most natural textures, however, exhibit a considerable degree of randomness in their structure, and it is therefore necessary to consider more than one texture element in order to detect the systematic geometric distortions due to the perspective effects. Attempts have been made at modeling such randomness statistically (Witkin 1981; Kanatani and Chou 1989; Blake and Marinos 1990a,b), but here such specific models are replaced by the basic principle of reducing variance by integration. For this reason, the concept of *texel grouping scale* has been introduced in the scheme above; it refers to the scale used for combining texture descriptors computed at different spatial points into entities to be used for computing geometric shape descriptors.

If the texture descriptors are combined by weighted averaging into a descriptor of the same type, as in the case of methods based on foreshortening, then the texel grouping scale is closely related (or even equivalent) to the relative integration scale. More precisely, from the semi-group property of Gaussian smoothing,  $g(\cdot; s_2) = g(\cdot; s_2 - s_1) * g(\cdot; s_1)$ , it follows that, if the local smoothing scale  $t$  is held constant, then the second moment matrix at any coarse integration scale,  $s_2$ , can be computed from the second moment matrices at any finer integration scale,  $s_1$ ,

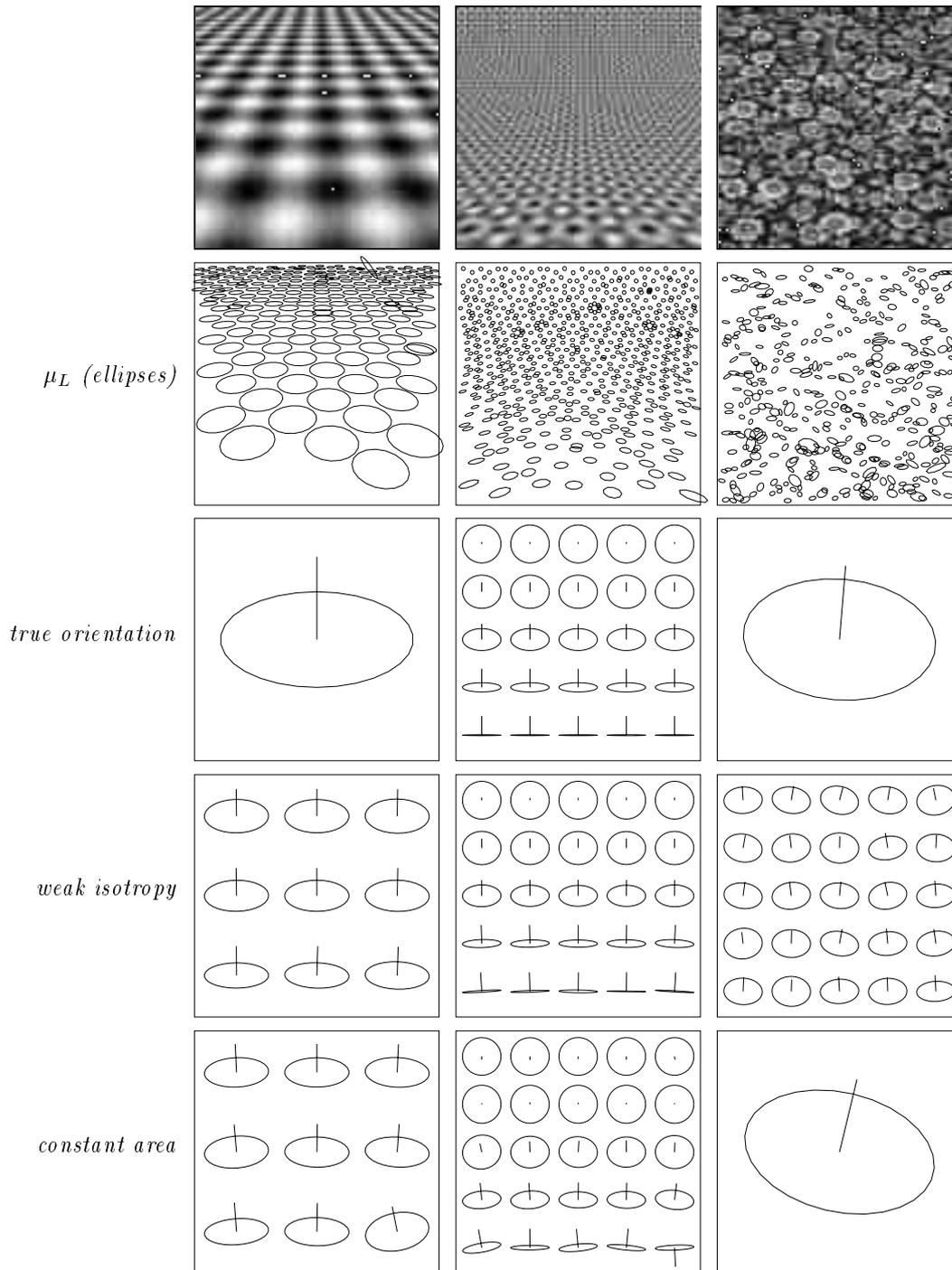
$$\mu_L(\cdot; t, s_2) = g(\cdot; s_2 - s_1) * \mu_L(\cdot; t, s_1). \quad (37)$$

Hence, if the local scale is constant (e.g. zero) then in the basic version of the method of estimating surface orientation from foreshortening and weak isotropy, the texel grouping scale is equivalent to the relative integration scale.

However, the cascade smoothing property (37) is not applicable when the texture descriptors are combined into a descriptor of a different type. For example, estimation of shape from texture gradients is based on the average rate of change of some property of the local texture descriptors, so in this case it is clearly not meaningful to compute an average texture descriptor for the whole region. Rather, the appropriate texture property (e.g. area) is estimated from each windowed second moment descriptor separately, and the corresponding texture gradient is then estimated using Gaussian weights given by the texel grouping scale. (The procedure for the case of the area gradient is described in Appendix A.2.)

So far no method for automatic selection of the texel grouping scale has been implemented. In the experiments presented below the estimates are computed on sparse regular grids, and the size of the Gaussian grouping window is proportional to the grid cells. An alternative approach is, of course, to let the texel grouping scale be proportional to the selected integration scale.

## 5.6 Experimental results



**Figure 11:** Estimating local surface orientation in a synthetic image of a planar surface with 1.4% noise (left), a synthetic image of a cylindrical surface with 25% noise (middle), and a real image of a planar surface with known orientation (right). The rows show from top to bottom; (a) the grey-level image, (b) elliptical blobs detected by the adaptive multi-scale method, (c) reference surface orientation, (d) surface orientation estimated from foreshortening, (e) surface orientation estimated from the area gradient.

The examples shown in this section have been computed using the integration scales  $s = 2t$  in the spatial and scale selection process, and  $s = 4t$  in the computation of  $\mu_L$  (i.e.  $\gamma_1 = \sqrt{2}$ ,  $\gamma_2 = \sqrt{2}$ ).

Figure 11 shows results<sup>9</sup> from two noisy synthetic images and one real image, all with known camera geometry and surface orientation. From top to bottom, the rows show the grey-level image, the detected blobs, the true surface orientation, the surface orientation estimated from foreshortening (only the first of the two estimates is shown), and the surface orientation estimated from the area gradient.

The synthetic image in the left column (also shown in Figure 8) contains a planar surface pattern consisting of the sum of two sine waves and 1.4% additive white Gaussian noise. The orientation of the surface is  $(\sigma = 60^\circ, \theta = 90^\circ)$ . Foreshortening can in principle be computed from each individual blob, and the  $3 \times 3$  grid of estimates is very accurate; the estimate in the center is  $(\hat{\sigma} = 60.8^\circ, \hat{\theta} = 90.1^\circ)$ . Computation of the area gradient requires at least three blobs, and the estimates are noticeably less accurate near the bottom of the image where the blob density is very low, and the boundary effects increase. The estimate in the center is  $(\hat{\sigma} = 61.6^\circ, \hat{\theta} = 89.0^\circ)$ .

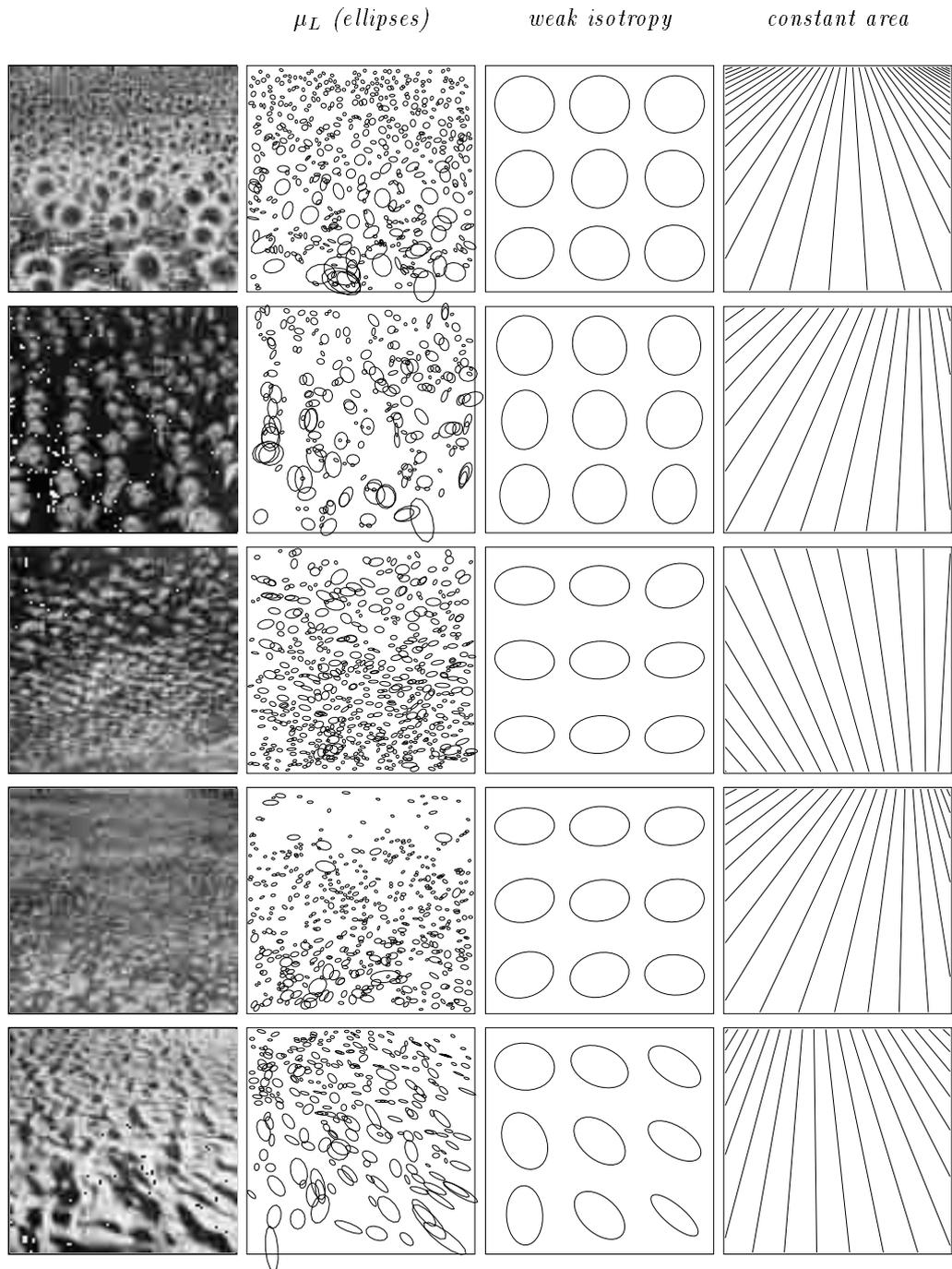
The middle column shows the same cylindrical surface image that was used in the first row in Figure 5. Here, 25% white Gaussian noise has been added; a noise level high enough to ensure that direct computations on unsmoothed data are bound to fail (compare with Table 1). It is quite obvious that the adaptive multi-scale blob detection technique is able to handle this noise level without much difficulty. At the center the true orientation is  $(\sigma = 55^\circ, \theta = 90^\circ)$ , the estimate from foreshortening is  $(\hat{\sigma} = 56.8^\circ, \hat{\theta} = 90.3^\circ)$ , and the estimate from the area gradient is  $(\hat{\sigma} = 36.1^\circ, \hat{\theta} = 86.2^\circ)$ . The fact that the slant of this surface is underestimated by the area gradient is entirely in keeping with the theory; the estimate is based on the assumption that  $\kappa_t = 0$  in (36), but here  $\kappa_t < 0$  since the surface is concave rather than flat.

The right column in Figure 11 shows the results obtained with a real image of planar surface with known surface orientation. The true surface orientation is  $(\sigma = 50.8^\circ, \theta = 85.3^\circ)$ , and at the center the estimate from foreshortening is  $(\hat{\sigma} = 50.1^\circ, \hat{\theta} = 85.0^\circ)$ . Considering the weak perspective effects resulting from the combination of a narrow field of view and a moderate slant, the estimate  $(\hat{\sigma} = 50.9^\circ, \hat{\theta} = 78.7^\circ)$  obtained from the area gradient in the whole image is surprisingly good. This may only have been a lucky coincidence, however; on a  $3 \times 3$  grid the estimates break down completely.

Figure 12 shows the results obtained with five images from (Blostein and Ahuja 1989). The camera geometry is unknown, and it is therefore impossible to compute absolute estimates of the surface orientation. To estimate surface orientation from foreshortening, the second moment matrix  $\mu_L(p)$  must first be transformed to  $T_p(\Sigma)$ , but the parameters of this transformation depend on the camera geometry and are hence unknown. Foreshortening is therefore visualized directly by ellipses representing

---

<sup>9</sup>In the examples in this section the surface orientation is indicated graphically by a dish with an attached needle parallel to the surface normal. In contrast to the previous illustrations of the second moment matrices, the dishes are from now on viewed in *parallel* projection along the visual ray through the image center. With this convention, the shape of each projected dish specifies the surface orientation regardless of the internal camera geometry and the position of the dish in the image.



**Figure 12:** Estimation of foreshortening and the area gradient in real images from (Blostein and Ahuja 1989). (a) Real grey-level image. (b) Elliptical blobs detected by the adaptive multi-scale method. (c) Estimated foreshortening, here represented by weighted averages of the second moment descriptors associated with each blob. (d) Estimated area gradient, visualized by lines aligned with the tilt direction converging to a point on the horizon.

the weighted second moment matrices in the image on which the estimate would be based. To estimate surface orientation from the area gradient, the focal length must be known. However, the position of the *horizon* of the plane, i.e., the line where projected area is estimated to vanish, can still be determined. The estimated horizon will often typically lie outside the image, but in the rightmost column of Figure 12 it is indirectly represented by a set of projected lines parallel to the tilt direction in the surface.

It is interesting to note that the foreshortening in these examples often reflects the orientation of the individual texture elements (e.g., the sunflowers), whereas the area gradient corresponds to the orientation of the underlying surface.

## 6 Shape from disparity gradients

In this section the application of the multi-scale second moment descriptor to shape estimation from binocular (stereo) vision will be briefly discussed.

Estimation of depth from stereo is based on the disparity cue, i.e., the slight difference in the left and right eyes' view of a point in the scene. If corresponding features in the left and right images can be identified, then depth can be recovered by triangulation provided that the viewing geometry is known.

The matching can be based on dense features, such as the local (possibly pre-filtered) brightness pattern itself (see e.g. (Barnard 1989)), or the projection of it onto some lower-dimensional set of basis functions (see e.g. (Jones and Malik 1992a)). Alternatively, it can be based on sparse but salient features such as edges or corners (see e.g. (Pollard et al 1985)). The latter approach typically reduces the computational cost, but has the drawback that it only produces isolated depth estimates, which have to be interpolated in order to find e.g. the local surface orientation.

It seems quite likely that the normalized scale-space extrema of various combinations of first- and second-order normalized Gaussian derivatives would be good candidates for pointwise matching. However, this approach will not be pursued here. Instead, we will focus on another aspect of the problem, namely, estimation of local surface shape when correspondence has already been established.

Briefly, the idea is to model the local transformation from the right eye's view of a small surface patch to the left eye's view of the same patch by an affine transformation rather than as a simple displacement. Establishing correspondence between a point in the left image with a point in the right image determines the translational part (and thereby distance if the viewing geometry is known). If, in addition, the linear part of the transformation can be estimated, then it is possible to recover additional information either about surface structure in the neighbourhood of the matched point, or about the viewing geometry. This approach was first studied by Koenderink and van Doorn (1976), and has more recently been developed by Wildes (1991), and Jones and Malik (1992b).

Following the latter work, let the left and right image coordinates of corresponding points be denoted  $(x_l, y_l)$  and  $(x_r, y_r)$ , respectively. The mapping  $M$  from  $(x_r, y_r)$  to  $(x_l, y_l)$  depends on the scene and viewing geometry, and it is in general non-linear. A

Taylor expansion of the first terms can be written

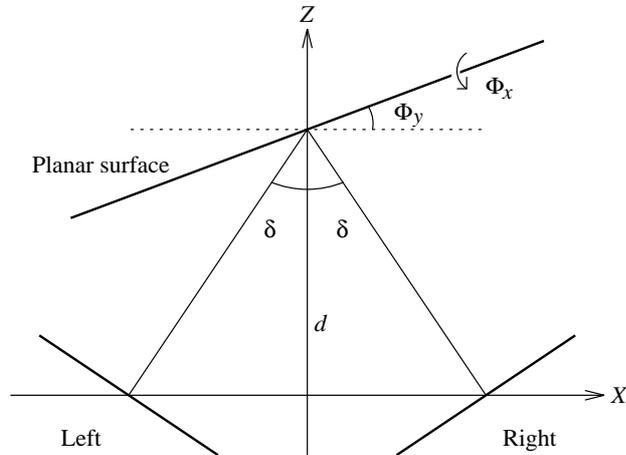
$$\begin{pmatrix} x_l \\ y_l \end{pmatrix} \approx \begin{pmatrix} 1 + H_x & H_y \\ V_x & 1 + V_y \end{pmatrix} \begin{pmatrix} x_r \\ y_r \end{pmatrix} + \begin{pmatrix} H \\ V \end{pmatrix}. \quad (38)$$

$(H_x, H_y)$  and  $(V_x, V_y)$  are often referred to as the gradients of horizontal and vertical disparity, respectively. In the following it will be assumed that the disparity  $(H, V)$  at the studied point is known, and without loss of generality we can then set  $H = V = 0$ .

The significance of (38) lies in the fact that if the parameters  $(H_x, H_y, V_x, V_y)$  can be estimated from the local structure of the brightness pattern in both images, they can then be interpreted in terms of surface and/or viewing geometry. Jones and Malik (1992b) estimated the parameters by trying a fixed number of transformations and selecting that which provided the best left-right fit. Here it will be shown that such an exhaustive search procedure is in fact unnecessary; the transformation parameters can be recovered directly from suitable local texture descriptors. First, however, the geometry of binocular perspective will be briefly reviewed.

## 6.1 Review of stereo geometry

The representation of the basic geometry shown in Figure 13 is based on (Jones and Malik 1992b). We also follow these authors in restricting the analysis to the fixation point and assuming symmetric vergence. The extension to asymmetric vergence and peripheral points is relatively straightforward, but the resulting algebraic expressions become more complicated.



**Figure 13:** Representation of view and surface geometry in the case of symmetric vergence.

The position and orientation of the planar surface relative to the cyclopean image plane  $Z = 0$  is represented by a rotation  $\phi_x$  around the  $x$  axis, followed by a rotation  $\phi_y$  around the  $y$  axis, and finally a translation  $d$  along the  $Z$  axis. The surface orientation relative to the left and right eyes' views are then  $(\phi_x, \phi_y + \delta)$  and  $(\phi_x, \phi_y - \delta)$ , respectively. The corresponding surface normal is  $(-\cos \phi_x \sin \phi_y, \sin \phi_x, \cos \phi_x \cos \phi_y)^T$ , and in terms of the slant-tilt representation we have

$$\cos \sigma = \cos \phi_x \cos \phi_y \quad \text{and} \quad \tan \theta = -\tan \phi_x / \sin \phi_y. \quad (39)$$

Let the image coordinates  $(x_l, y_l)$  and  $(x_r, y_r)$  be defined with respect to the image of the fixation point. To first order, the mapping from  $(x_r, y_r)$  to  $(x_l, y_l)$  in the neighbourhood of the fixation point is then

$$\begin{pmatrix} \frac{\cos(\phi_y - \delta)}{\cos(\phi_y + \delta)} & -\tan \phi_x \frac{\sin 2\delta}{\cos(\phi_y + \delta)} \\ 0 & 1 \end{pmatrix} \quad (40)$$

(see (Jones and Malik 1992b) for a derivation). Hence, under these viewing conditions, the gradient of vertical disparity is identically zero. Comparing with (38), we have

$$\tan \phi_y = \frac{H_x}{H_x + 2} \cot \delta, \quad (41)$$

$$\tan \phi_x = -\frac{H_y}{\sqrt{H_x^2 + 4(H_x + 1) \sin^2 \delta}}, \quad (42)$$

which shows that the gradient of horizontal disparity uniquely determines the surface orientation.

## 6.2 Estimating disparity gradients

In principle, the disparity gradient  $(H_x, H_y)$  can be estimated by differentiation of the disparity obtained from a dense set of point-to-point matches. However, obtaining a dense disparity map is a costly and sometimes unreliable process. We therefore take an alternative approach, based on matching only a *single* left-right point pair, and then analyzing the structure of the brightness pattern in the neighbourhood of both image points.

Let  $\mu_L$  and  $\mu_R$  denote the windowed second moment matrices computed at the left and right images of the fixation point. If the linearized mapping from the left image to the right is denoted  $M_*$ , then from (7)

$$\mu_L = M_*^T \mu_R M_*. \quad (43)$$

If  $\mu_L$  and  $\mu_R$  are known, then (43) provides three equations for the four parameters of the unknown linear transformation  $M_*$ . Hence, to fully recover  $M_*$  it is necessary to provide one additional constraint. The viewing geometry described above in fact provides *two* additional constraints, since  $V_x = V_y = 0$ . Hence, the relation (40) can be expressed as

$$M_* = \begin{pmatrix} 1 + H_x & H_y \\ 0 & 1 \end{pmatrix}, \quad (44)$$

which when substituted into (43) yields

$$\mu_L = \begin{pmatrix} (1 + H_x)^2 \mu_{11}^R & (1 + H_x)(H_y \mu_{11}^R + \mu_{12}^R) \\ (1 + H_x)(H_y \mu_{11}^R + \mu_{12}^R) & H_y^2 \mu_{11}^R + 2H_y \mu_{12}^R + \mu_{22}^R \end{pmatrix}, \quad (45)$$

where  $\mu_{ij}^R$  denotes the components of  $\mu_R$ .

Hence, in this case the problem of estimating  $M_*$  from the left and right texture descriptors  $\mu_L$  and  $\mu_R$  is overdetermined. An obvious possibility is to estimate  $(H_x, H_y)$  by a least squares fit of (45). This is, however, not necessarily the most useful approach. With reference to the ellipse representation of a second moment matrix, a more interesting possibility is to compute  $(H_x, H_y)$  from the difference in *shape* of  $\mu_L$  and  $\mu_R$ , while ignoring any difference in *size*. This idea can be motivated by the fact that the estimated shape is typically more reliable than the estimated size (see Section 3.6). Moreover, it has interesting connections with certain theories of human stereo vision, as discussed below.

The shape of a second moment matrix  $\mu$  is represented by the normalized components  $(\tilde{C}, \tilde{S})^T$  defined by (9) and (12). Substituting these relations into (45), we obtain after some algebraic manipulation

$$H_x = \frac{1 + \tilde{C}_L}{1 + \tilde{C}_R} \sqrt{\frac{1 - \tilde{Q}_R^2}{1 - \tilde{Q}_L^2}} - 1, \quad (46)$$

$$H_y = \frac{\tilde{S}_L \sqrt{1 - \tilde{Q}_R^2} - \tilde{S}_R \sqrt{1 - \tilde{Q}_L^2}}{(1 + \tilde{C}_R) \sqrt{1 - \tilde{Q}_L^2}}. \quad (47)$$

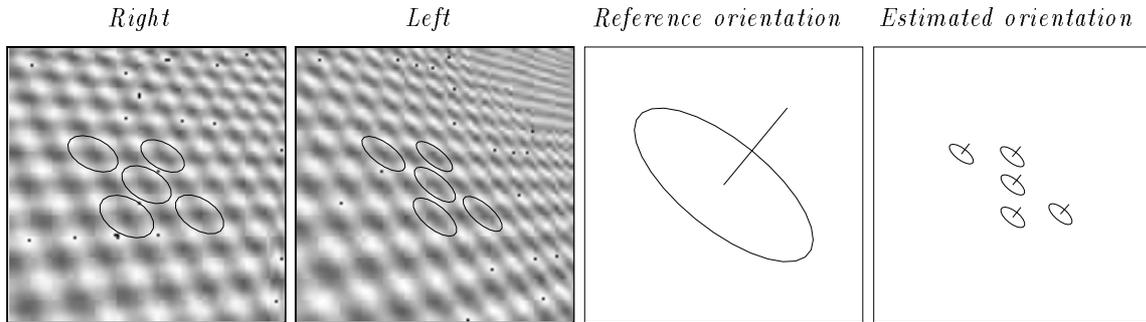
To summarize, (46) and (47) express the gradient of horizontal disparity in closed form, in terms of the local normalized directional statistics of first-order derivatives of the image brightness in corresponding patches of the left and right images. The expressions are invariant with respect to uniform scaling of the brightness of one image with respect to the other, and, more importantly, also to uniform magnification of one image with respect to the other.

This method is closely related, but not equivalent, to the method proposed by Koenderink and van Doorn (1976) based on the *def* component of the disparity gradient; the methods coincide only when the disparity gradient is infinitesimally small. It is interesting to note that both methods account at least qualitatively for the so-called “induced effect” (see e.g. (Mayhew and Longuet-Higgins 1982)), i.e., the phenomenon that vertical magnification of one eye’s view affects the slant perceived by a human observer, despite the fact that all horizontal disparities remain unchanged.

### 6.3 Experimental results

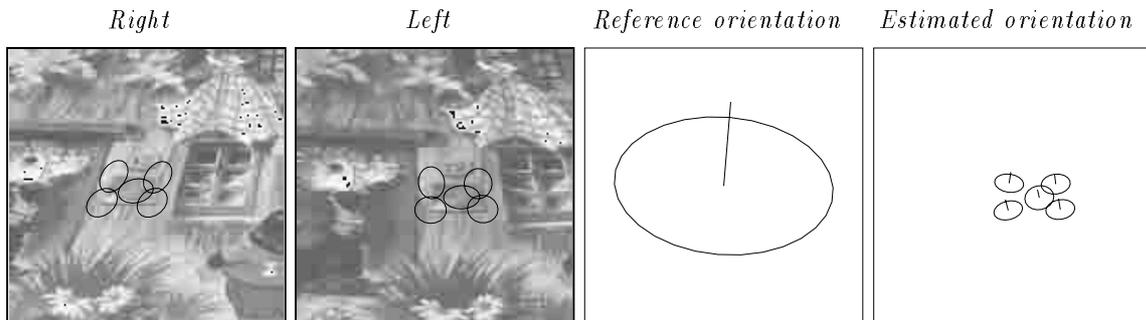
Figure 14 shows the ellipse representation of the window second moment matrix computed at the fixation point and four neighbouring points (using  $\gamma_1 = \sqrt{2}, \gamma_2 = 2\sqrt{2}$ ) superimposed on a bright copy of a synthetic stereo pair (arranged for cross-eyed fusion). The images are perspective views of a sinusoidal pattern, and contain 5% additive Gaussian noise. The visual angle across the diagonal of each image is  $32^\circ$ , and the vergence angle is  $\delta = 10^\circ$ . The orientation of the surface is  $(\phi_x = -45^\circ, \phi_y = 45^\circ)$ , corresponding to  $(\sigma = 60^\circ, \theta = 54.7^\circ)$  in the slant-tilt representation.

At the fixation point, the estimated disparity gradient was  $(\hat{H}_x = 0.405, \hat{H}_y = 0.577)$ , and from (41) and (42) we then obtain the estimated surface orientation  $(\hat{\phi}_x = -45.0^\circ, \hat{\phi}_y = 43.7^\circ)$ , or in terms of slant and tilt  $(\hat{\sigma} = 59.2^\circ, \hat{\theta} = 55.3^\circ)$ . The



**Figure 14:** Local surface orientation estimated from the gradient of horizontal disparity in a synthetic stereo pair with 5% noise. The columns show from left to right; (a-b) Bright copies of the right and left images with the computed texture descriptor superimposed. (c) reference surface orientation, (d) estimated surface orientation at five manually matched points. (c) and (d) are shown with respect to the left eye’s view.

error in the estimate, expressed as the angle between the estimated and true surface normals, is only  $0.9^\circ$ . The results obtained at the remaining four points were very similar, as is evident from the graphical representation show to the right in Figure 14.



**Figure 15:** Local surface orientation estimated from the gradient of horizontal disparity in a real stereo pair. The columns show from left to right; (a-b) Bright copies of the right and left images with the computed texture descriptor superimposed. (c) reference surface orientation, (d) estimated surface orientation at five manually matched points. (c) and (d) are shown with respect to the left eye’s view.

Figure 15 shows the results obtained by applying the same procedure to a real stereo pair. Five point pairs were matched manually, and  $\mu_L$  and  $\mu_R$  were then computed at each of these points, using the integration scale determined at the central point to simplify comparison of the results. Surface orientation was then estimated separately for each point pair. The results are displayed to the right in Figure 15. There appears to be a slight underestimation of the local anisotropy in both images, and for this surface orientation where  $|\phi_y| \ll |\phi_x|$  these errors are added rather than cancelled, leading to a systematic bias to the left in the estimated tilt. We are currently investigating in more detail how the stability of the method depends on the true surface orientation.

## 7 Summary and discussion

We have shown that a representation of local image structure computed by multi-scale bottom-up retinotopic processing can be directly used to derive non-trivial cues to the local structure of three-dimensional surfaces in the scene, without iterations, search, or high-level knowledge.

In the first part of the paper, we treated the problem of computing such a representation, and introduced the windowed second moment matrix to represent the local statistics of first order Gaussian normalized derivatives of image brightness. We showed that linear transformations of the spatial coordinates affect this descriptor in a simple way, which allows the parameters of the transformation to be estimated from the properties of the descriptor.

The computation of this descriptor involves two scale parameters; first, the smoothing scale at which derivatives of the image brightness are computed, and second, the scale of the window used to integrate statistics of nonlinear descriptors of the differential image structure. We proposed a systematic two-stage method for adaptively choosing these scale parameters. The characteristic dimensions of salient image structures at any given point are first estimated by detecting local maxima with respect to scale of certain differential invariants derived from the windowed second moment matrix. The integration scale is then set proportional to the estimated characteristic dimensions, while the smoothing scale is adapted to obtain a trade-off between suppression of noise and irrelevant fine-scale structures on the one hand, and distortion of the shape of local image structures due to smoothing on the other.

The principle used to determine characteristic dimension was also applied to guide the selection of where in the image to compute the texture descriptors. Whereas the second moment descriptor which describes local image “shape” is based on first derivatives, the entities used for spatial selection were based on second derivatives in order to favour centers of blob-like structures.

In the second part of the paper we treated the problem of using the multi-scale second moment descriptor to derive cues to local three-dimensional surface shape and orientation. We first discussed estimation of shape from texture in a monocular image, based on two independent cues referred to as foreshortening and the area gradient, respectively. It was shown that these two cues can be reliably computed both in noisy synthetic images and natural images.

We then showed that the same methodology can be used to recover local surface orientation by estimating the gradient of horizontal disparity in a binocular image pair. This method has the advantage that it does not depend on any assumptions about the surface texture. We presented experimental results obtained by applying the method to synthetic and natural images.

### 7.1 Relations to biological vision

As mentioned in the introduction, we have not attempted to model biological vision on a detailed level. However, the general principles on which the methodology is based appear to be compatible with current understanding of the structure of the first stages of the primate visual pathway.

For example, it is worth noting that the computation of the windowed second moment descriptor follows the pattern “linear filtering – nonlinearity – spatial averaging”. Processing sequences of this type have in recent years been proposed as models for human texture discrimination, e.g. (Caelli 1985; Bergen and Adelson 1988; Malik and Perona 1990). The initial linear filtering stage in our model is based on directional Gaussian derivatives, which have been used to model the receptive fields of simple cells in the mammalian visual cortex (Young 1985). Moreover, selection of scale levels and spatial positions by detection of local maxima could easily be implemented by lateral inhibition between cells.

The spatial detection process we discussed was based on rotationally symmetric operators such as the Laplacian, which limits the ability to detect very elongated blob-like structures based on the response of a single operator. However, in this context it is interesting to note that in a psychophysical study of the visibility of elliptical Gaussian blobs, Bijl and Koenderink (1993) found that their results can be predicted by a model based on Pythagorean summation of the responses of rotationally symmetric receptive fields.

## 7.2 Further research

Some issues not directly addressed by the present work are discussed below.

**Grouping** We have tacitly assumed that integration of local properties is always a meaningful operation, but in general situations it may be necessary to restrict the integration to some coherent subset of the descriptors in the window. This can have any of a number of reasons, e.g. that the image contains more than one surface, that a surface contains more than one type of texture, or that an image region contains textures resulting from more than one physical process.

Furthermore, we have in most cases used only the most dominant scale at each spatial position; a more general approach would be to detect all local maxima, and then apply spatial grouping based on similarity of characteristic dimension. For example, a noisy image of a slanted pattern might give rise to maxima at small scales due to the noise, in addition to the maxima at coarser scales corresponding to the surface texture. Separate estimation of the area gradient for the fine-scale maxima would then correctly indicate a fronto-parallel surface corresponding to the noise in the image plane.

**Cue combination** This paper has treated local estimation of surface shape and orientation, using three independent processes. Clearly, some mechanism is needed for unifying these independent estimates into hypotheses about coherent surfaces.

**Brightness discontinuities** The linear transformation property (7) is strictly valid only if the brightness pattern is differentiable. Non-differentiable structures such as sharp discontinuities may therefore invalidate (7) to a greater or lesser extent. For example, compression of an ideal step edge in the direction perpendicular to the edge obviously does not affect the magnitude of derivatives estimated by finite differences

at all, unlike the case of a smooth edge for which the compression would affect the slope of the edge. We plan to investigate this problem in more detail.

**Non-uniform smoothing** The reason for adapting the local scale in the computation of the second moment descriptor was to obtain a reasonable trade-off between on the one hand suppression of noise and irrelevant fine-scale structures, and on the other hand distortion of the shape of the brightness pattern due to the isotropic Gaussian smoothing.

However, if the shape of the smoothing kernel is adapted to have the same anisotropy as the brightness pattern, then the shape distortion effect is reduced (Lindeberg and Gårding 1993). This observation is related to the suggestion by Stone (1990) to adapt the local operators used in shape-from-texture estimation to be isotropic when backprojected to the surface, rather than in the image.

## A Appendix

### A.1 Transformation property of the second moment matrix

The transformation property (7) of the windowed second moment matrix can be verified as follows. Assume that  $L, R : \mathbb{R}^2 \rightarrow \mathbb{R}$  are two intensity patterns related by  $L(\xi) = R(B\xi)$ , where  $\xi \in \mathbb{R}^2$ , and  $B$  is a non-singular linear transformation. Without loss of generality assume  $\det B > 0$ . Then,

$$\nabla L(\xi) = B^T \nabla R(B\xi),$$

which when substituted into the definition of the windowed second moment matrix yields

$$\mu_L(q) = \iint_{\xi \in \mathbb{R}^2} w(q-\xi) (\nabla L(\xi)) (\nabla L(\xi))^T d\xi = \iint_{\xi \in \mathbb{R}^2} w(q-\xi) B^T (\nabla R(B\xi)) (\nabla R(B\xi))^T B d\xi.$$

Substituting  $\eta = B\xi$  (with  $p = Bq$ ) we obtain

$$\mu_L(q) = B^T \left\{ \iint_{\eta \in \mathbb{R}^2} w(B^{-1}(p-\eta)) (\nabla R(\eta)) (\nabla R(\eta))^T (\det B)^{-1} d\eta \right\} B.$$

The integral within brackets is the second moment of  $R$  at  $p$  computed with respect to the backprojected window function  $w'(\eta-p) = (\det B)^{-1} w(B^{-1}(\eta-p))$ . This window function is normalized as long as the original window function is, because

$$\iint_{\eta \in \mathbb{R}^2} w(B^{-1}(\eta-p)) (\det B)^{-1} d\eta = \{\text{let } \eta = B\xi \text{ with } p = Bq\} = \iint_{\xi \in \mathbb{R}^2} w(\xi-q) d\xi,$$

which verifies (7). Note, however, that the window function  $w'$  will not, in general, be rotationally symmetric.

## A.2 Estimating simple distortion gradients

In this appendix a practical procedure for estimation of surface orientation from the area gradient in the case of a locally planar surface will be described. A more detailed description is given in (Lindeberg and Gårding 1993). The same procedure can with only minor modifications be applied to estimation of surface orientation from any simple distortion gradient.

Equation (36) relates the normalized gradient of projected texel area in the view-sphere  $\Sigma$  to surface orientation and curvature. If the curvature is assumed to be small, an estimate of the surface tilt is given by the negative direction of the area gradient, and an estimate of surface slant is given by  $\tan^{-1} \|(\nabla A_\Sigma(p))/(3A_\Sigma(p))\|$ .

In principle, the area gradient can be estimated by applying a central difference operator to the product  $mM$  obtained from the pointwise estimate of  $F_*$ . However, for a planar surface the product  $A_\Sigma = mM$  is not a linear function of the image coordinates, and so a central difference estimate of the first derivative would be biased by the higher derivatives of  $A_\Sigma$ . A more consistent approach is to transform  $A_\Sigma$  to a form that is linear in the image before the central difference operator is applied, thereby eliminating the bias. This procedure can be simplified even further by transforming the image texel area  $A_L$ , rather than the viewsphere texel area  $A_\Sigma$ , to linear form, thus bypassing the need to apply the gaze transformation  $G_*$ .

For a planar surface with slant  $\sigma$  and tilt  $\theta$ , it can be shown (Lindeberg and Gårding 1993) that

$$(A_L(x, y))^{1/3} = k(f \cos \sigma - (x \cos \theta + y \sin \theta) \sin \sigma), \quad (48)$$

where  $k$  is an unknown constant.<sup>10</sup>

Hence, a practical procedure for estimating the local surface orientation from estimates of, for example, the area  $A_L(x, y)$  in some region of the image can be described as follows. First, compute (samples of)  $h(x, y) = (A_L(x, y))^{1/3}$ . Then, estimate the parameters  $(h_x, h_y, h(0, 0))$ , either by central differences or, more robustly, by a weighted least-squares fit of  $h(x, y) = h_x x + h_y y + h(0, 0)$ . Finally, compute the estimated local surface orientation using (48) which can be rewritten

$$\hat{\sigma} = \cos^{-1} \left( \frac{h(0, 0)}{\sqrt{f^2 h_x^2 + f^2 h_y^2 + h^2(0, 0)}} \right), \quad (49)$$

$$\hat{\theta} = \arg(h_x, h_y). \quad (50)$$

Note that this procedure only requires  $A_L(x, y)$  to be computed up to an arbitrary scale factor.

## References

- [1] Y. Aloimonos, "Shape from texture", *Biological Cybernetics*, vol. 58, pp. 345–360, 1988.

<sup>10</sup>Similar expressions have been derived e.g. by Blostein and Ahuja (1989), and Kanatani and Chou (1989).

- [2] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, “Uniqueness of the Gaussian kernel for scale-space filtering”, *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 8, no. 1, pp. 26–33, 1986.
- [3] S. T. Barnard, “Stochastic stereo matching over scale”, *Int. J. of Computer Vision*, vol. 3, pp. 17–22, 1989.
- [4] J. R. Bergen and E. H. Adelson, “Early vision and texture perception”, *Nature*, vol. 333, pp. 363–364, 1988.
- [5] J. Bigün, G. H. Granlund, and J. Wiklund, “Multidimensional orientation estimation with applications to texture analysis and optical flow”, *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 13, pp. 775–790, Aug. 1991.
- [6] P. Bijl and J. J. Koenderink, “Visibility of elliptical Gaussian blobs”, *Vision Research*, vol. 33, no. 2, pp. 243–255, 1993.
- [7] A. Blake and C. Marinos, “Shape from texture: estimation, isotropy and moments”, *J. of Artificial Intelligence*, vol. 45, pp. 323–380, 1990.
- [8] A. Blake and C. Marinos, “Shape from texture: the homogeneity hypothesis”, in *Proc. 3rd Int. Conf. on Computer Vision*, (Osaka, Japan), pp. 350–353, IEEE Computer Society Press, Dec. 1990.
- [9] D. Blostein and N. Ahuja, “A multiscale region detector”, *Computer Vision, Graphics, and Image Processing*, vol. 45, pp. 22–41, 1989.
- [10] D. Blostein and N. Ahuja, “Shape from texture: integrating texture element extraction and surface estimation”, *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 11, pp. 1233–1251, Dec. 1989.
- [11] L. G. Brown and H. Shvaytser, “Surface orientation from projective foreshortening of isotropic texture autocorrelation”, *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 12, pp. 584–588, June 1990.
- [12] T. Caelli, “Three processing characteristics of visual texture segmentation”, *Spatial Vision*, vol. 1, pp. 19–30, 1985.
- [13] L. S. Davis, L. Janos, and S. M. Dunn, “Efficient recovery of shape from texture”, *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 5, pp. 485–492, Sept. 1983.
- [14] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, “Scale and the differential structure of images”, *Image and Vision Computing*, vol. 10, pp. 376–388, July/August 1992.
- [15] M.A. Förstner and E. Gülch, “A fast operator for detection and precise location of distinct points, corners and centers of circular features”, in *ISPRS Intercommission Workshop*, 1987.
- [16] J. Gårding, *Shape from surface markings*. PhD thesis, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, May 1991.
- [17] J. Gårding, “Shape from texture for smooth curved surfaces in perspective projection”, *J. of Mathematical Imaging and Vision*, vol. 2, pp. 329–352, 1992.
- [18] J. Gårding, “Shape from texture and contour by weak isotropy”, *J. of Artificial Intelligence*, 1993. (In press).
- [19] J. Gibson, *The Perception of the Visual World*. Houghton Mifflin, Boston, 1950.

- [20] D. G. Jones and J. Malik, "A computational framework for determining stereo correspondences from a set of linear spatial filters", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, pp. 395–410, Springer-Verlag, May 1992.
- [21] D. G. Jones and J. Malik, "Determining three-dimensional shape from orientation and spatial frequency disparities", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, pp. 661–669, Springer-Verlag, May 1992.
- [22] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex", *J. of Neurophysiology*, vol. 58, pp. 1233–1258, 1987.
- [23] J. Jones and L. Palmer, "The two-dimensional spatial structure of simple receptive fields in cat striate cortex", *J. of Neurophysiology*, vol. 58, pp. 1187–1211, 1987.
- [24] B. Julesz, "Textons, the elements of perception and their interactions", *Nature*, vol. 290, pp. 91–97, 1981.
- [25] K. Kanatani, "Detection of surface orientation and motion from texture by a stereological technique", *J. of Artificial Intelligence*, vol. 23, pp. 213–237, 1984.
- [26] K. Kanatani and T. C. Chou, "Shape from texture: general principle", *J. of Artificial Intelligence*, vol. 38, pp. 1–48, 1989.
- [27] J. J. Koenderink, "The structure of images", *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.
- [28] J. J. Koenderink and A. J. van Doorn, "Geometry of binocular vision and a model for stereopsis", *Biological Cybernetics*, vol. 21, pp. 29–35, 1976.
- [29] J. J. Koenderink and A. J. van Doorn, "Receptive field families", *Biological Cybernetics*, vol. 63, pp. 291–298, 1990.
- [30] T. Lindeberg, "Scale-space for discrete signals", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 12, pp. 234–254, Mar. 1990.
- [31] T. Lindeberg, *Discrete scale space theory and the scale space primal sketch*. PhD thesis, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, May 1991. A revised and extended version to appear in Kluwer Int. Series in Engineering and Computer Science.
- [32] T. Lindeberg, "Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction", *J. of Mathematical Imaging and Vision*, Apr. 1992. (In press).
- [33] T. Lindeberg, "On scale selection for differential operators", in *Proc. 8th Scandinavian Conf. on Image Analysis*, (Tromsø, Norway), May 1993. To appear.
- [34] T. Lindeberg and J.O. Eklundh, "The scale-space primal sketch: Construction and experiments", *Image and Vision Computing*, vol. 10, pp. 3–18, Jan. 1992.
- [35] T. Lindeberg and J. Gårding, "Shape from texture from a multi-scale perspective", Tech. Rep. CVAP116, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Jan. 1993. A shortened version to appear at 4th ICCV, (Berlin, Germany), May, 1993.
- [36] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms", *J. of the Optical Society of America*, vol. 7, pp. 923–932, 1990.

- [37] K. V. Mardia, *Statistics of Directional Data*. Academic Press, London, 1972.
- [38] D. Marr, *Vision*. W.H. Freeman, New York, 1982.
- [39] D. C. Marr, “Early processing of visual information”, *Phil. Trans. Royal Soc (B)*, vol. 27S, pp. 483–524, 1976.
- [40] J. E. W. Mayhew and H. C. Longuet-Higgins, “A computational model of binocular depth perception”, *Nature*, vol. 297, pp. 376–378, 1982.
- [41] B. O’Neill, *Elementary Differential Geometry*. Academic Press, Orlando, Florida, 1966.
- [42] A. P. Pentland, “Shading into texture”, *J. of Artificial Intelligence*, vol. 29, pp. 147–170, 1986.
- [43] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, “PMF: A stereo correspondence algorithm using a disparity gradient limit”, *Perception*, vol. 14, pp. 449–470, 1985.
- [44] A. R. Rao and B. G. Sunk, “Computing oriented texture fields”, *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 157–185, Mar. 1991.
- [45] J. V. Stone, “Shape from texture: textural invariance and the problem of scale in perspective images of surfaces”, in *Proc. British Machine Vision Conference*, (Oxford, England), Sept. 1990.
- [46] B. J. Super and A. C. Bovik, “Shape-from-texture by wavelet-based measurement of local spectral moments”, in *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, (Champaign, Illinois), pp. 296–301, June 1992.
- [47] M. R. Turner, “Texture discrimination by Gabor functions”, *Biological Cybernetics*, vol. 55, pp. 71–82, 1986.
- [48] H. Voorhees and T. Poggio, “Detecting textons and texture boundaries in natural images”, in *Proc. 1st Int. Conf. on Computer Vision*, (London, England), 1987.
- [49] R. P. Wildes, “Direct recovery of three-dimensional scene geometry from binocular stereo disparity”, *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 13, no. 8, pp. 761–774, 1981.
- [50] A. P. Witkin, “Recovering surface shape and orientation from texture”, *J. of Artificial Intelligence*, vol. 17, pp. 17–45, 1981.
- [51] A. P. Witkin, “Scale-space filtering”, in *Proc. 8th Int. Joint Conf. Art. Intell.*, (Karlsruhe, West Germany), pp. 1019–1022, Aug. 1983.
- [52] R. A. Young, “The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles”, Tech. Rep. GMR-4920, Computer Science Department, General Motors Research Lab., Warren, Michigan, 1985.
- [53] R. A. Young, “The Gaussian derivative model for spatial vision: I. Retinal mechanisms”, *Spatial Vision*, vol. 2, pp. 273–293, 1987.
- [54] A. L. Yuille and T. A. Poggio, “Scaling theorems for zero-crossings”, *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 8, pp. 15–25, 1986.