

# Shape from Texture from a Multi-Scale Perspective

Tony Lindeberg and Jonas Gårding

Computational Vision and Active Perception Laboratory (CVAP)\*  
Royal Institute of Technology (KTH), Stockholm, Sweden

**Abstract:** The problem of *scale* in shape from texture is addressed. The need for (at least) two scale parameters is emphasized; a *local scale* describing the amount of smoothing used for suppressing noise and irrelevant details when computing primitive texture descriptors from image data, and an *integration scale* describing the size of the region in space over which the statistics of the local descriptors is accumulated.

A novel mechanism for automatic *scale selection* is proposed, based on *normalized derivatives*. It is used for *adaptive* determination of the two scale parameters in a multi-scale texture descriptor, the *windowed second moment matrix*, which is defined in terms of Gaussian smoothing, first order derivatives, and non-linear pointwise combinations of these. The same scale-selection method can be used for multi-scale blob detection without any tuning parameters or thresholding.

The resulting texture description can be combined with various assumptions about surface texture in order to estimate local surface orientation. Two specific assumptions, “weak isotropy” and “constant area”, are explored in more detail. Experiments on real and synthetic reference data with known geometry demonstrate the viability of the approach.

## 1 Introduction

This paper addresses the problem of *scale* in shape-from-texture. Any shape-from-texture method, e.g. [14, 35, 29, 1, 7, 15, 8, 5, 13], must begin by extracting some description of the image texture. In real-life situations it is hardly ever possible to know in advance at what scales to compute these. Size variations of image structures can occur because a surface texture contains structures of different physical size, because of artifacts in the image formation process, and also because of the very perspective effects one is trying to analyze.

---

\*We would like to thank Narendra Ahuja and John P. Frisby for kindly providing some of the images used in the paper. This work was partially performed under the ESPRIT-BRA project INSIGHT, and with support from the Swedish National Board for Industrial and Technical Development, NUTEK. The second author is also grateful for the support of the Swedish Institute and the Foundation Blanceflor Boncompagni-Ludovisi, née Bildt. Address: NADA, KTH, S-100 44 Stockholm, Sweden  
Email: tony@bion.kth.se, jonasg@bion.kth.se

Although (traditional) *scale-space theory* [36, 17, 2, 37, 20, 18, 10, 23] provides a methodology for handling such size or scale variations in image data, it does not address the problem how to *select* appropriate scales and structures for further analysis. As a new tool for approaching this problem, we propose a heuristic principle that *local extrema over scale* of different (possibly non-linear) combinations of *normalized scale-invariant derivatives* are likely to correspond to interesting structures. We also emphasize the need for and make explicit use of two different scale parameters in the shape from texture problem; a *local scale* describing the amount of smoothing used for suppressing irrelevant fine scale structures when computing non-linear descriptors from grey-level data, and an *integration scale* describing the size of the region in space used for accumulating the statistics of such local descriptors into entities that can be used as cues to the structure of the 3D world.

The proposed computational model is expressed completely in terms of *differential invariants* defined from *Gaussian derivatives* at multiple scales in scale-space. This makes it attractive for theoretical analysis and immediate implementation in a *visual front end*. In fact, shape cues can be *directly* computed from image data using only the following types of retinotopic operations; (large support) diffusion smoothing, (small support) derivative computations, and (pointwise) non-linear combinations of these.

Due to lack of space the presentation is heavily condensed; see [25] for an extensive description.

## 2 Background: Shape from texture

This section reviews the shape-from-texture problem, and defines the multi-scale image descriptor we propose.

### 2.1 Review of imaging geometry

Following [12], consider the perspective mapping of a smooth surface  $S$  onto a unit viewsphere  $\Sigma$ . At any point  $p$  on  $\Sigma$  let  $(\bar{p}, \bar{t}, \bar{b})$  be a local orthonormal coordinate system with  $\bar{p}$  as view direction. The *tilt 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, and by  $F_{*p} : T_p(\Sigma) \rightarrow$

$T_{F(p)}(S)$  the derivative (linear approximation) of  $F$  at any point  $p$  on  $\Sigma$ , where  $T_p(\Sigma)$  is the tangent plane of  $\Sigma$  at  $p$ , and  $T_{F(p)}(S)$  is the tangent plane of  $S$  at  $F(p)$ . 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  $F_{*p}$  is

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

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. The inverse eigenvalues of  $F_{*p}$ ,  $m < M$ , 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  $\bar{t}$ ) and  $M$  as major axis (parallel to  $\bar{b}$ ).

If the local linearized transformation  $F_{*p}$  can be estimated (up to certain scale factors) from the local image structure, it can then be directly interpreted in terms of surface properties. For example, *foreshortening*  $m/M = \cos\sigma$  provides direct information about surface orientation, and the normalized *area gradient*  $\nabla(mM)/(mM)$  can be shown to encode both surface orientation and curvature.

When image data are given in a planar image  $\Pi$  rather than on the viewsphere  $\Sigma$ ,  $F_{*p}$  can nevertheless be computed from the derivative  $A_{*q} = F_{*p}G_{*q}$  of the composed mapping  $A = F \circ G$ , since the derivative of  $G : \Pi, q \rightarrow \Sigma, p$  is known if the camera geometry is.

## 2.2 Definition of image texture descriptor

Let  $L(x, y)$  be the image brightness, and let  $\nabla L = (L_x, L_y)^T$  be its gradient. Given a symmetric and normalized window function  $w$ , a useful image texture descriptor is *the windowed second moment matrix* [25],

$$\mu_L(q) = \int_{x \in \mathbb{R}^2} (\nabla L(x))(\nabla L(x))^T w(q-x) dx, \quad (2)$$

where  $q \in \Pi$  denotes the image point where it is computed (compare with [11, 8, 5, 4, 31, 33, 12, 13]). From its components,  $\mu_{ij} = E_q(L_{x^i}L_{x^j})$  (where  $E_q$  represents the operator corresponding to (2)), introduce

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

as well as the two derived (anisotropy) measures

$$Q = \sqrt{C^2 + S^2}, \quad \tilde{Q} = Q/P. \quad (3)$$

It can be shown [25] that  $\mu_L(q)$  is a well-defined (coordinate independent) differential geometric entity in the image domain. It is invariant to translations, and has a nice behaviour with respect to uniform rescalings in the

spatial domain and affine brightness transformations. Although the matrix itself is not invariant to rotations, its eigenvectors,  $(\bar{e}_1, \bar{e}_2)$ , and eigenvalues,  $\lambda_1 \geq \lambda_2$ , are.

$$\lambda_{1,2} = P \pm Q = P(1 \pm \tilde{Q}). \quad (4)$$

The vector  $(C, S)^T$  is the *average* of all gradient vectors mapped to the double angle [27, 16] and weighted by the gradient magnitude squared multiplied by the window function. The direction of  $\arg(C, S)^T/2$  is parallel to  $\bar{e}_1$  and represents the *unsigned* average direction of the weighted gradient distribution. The *normalized anisotropy*  $\tilde{Q} \in [0, 1]$ ; it holds that  $\tilde{Q} = 0$  iff  $E_q(L_x^2) = E_q(L_y^2)$  and  $E_q(L_xL_y) = 0$ , while  $\tilde{Q} = 1$  iff  $(E_q(L_xL_y))^2 = E_q(L_x^2)E_q(L_y^2)$ . For example, a rotationally symmetric pattern has  $\tilde{Q} = 0$ , while  $\tilde{Q} = 1$  for a translationally symmetric distribution. Rotational symmetry is, however, not necessary in order to give  $\tilde{Q} = 0$ ; e.g., any pattern with  $N \geq 2$  uniformly distributed (unsigned) directions satisfies  $\tilde{Q} = 0$ .

## 2.3 Linear transformation property

Under an invertible linear transformation of the image domain  $\eta = B\xi$ , where  $B : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  and  $\xi, \eta \in \mathbb{R}^2$ , it can be shown that if a transformed intensity pattern  $R$  is defined by  $L(\xi) = R(B\xi)$ , then  $\mu_L(q)$  transforms as

$$\mu_L(q) = B^T \mu_R(p) B, \quad (5)$$

where  $\mu_R(p)$  is the second moment matrix of  $R$  at  $p = Bq$  computed with respect to the ‘‘backprojected’’ normalized window function  $w'(\eta-p) = (\det B)^{-1}w(\xi-q)$ .

## 2.4 Deriving shape cues from image data

The importance of the second moment matrix with respect to shape-from-texture analysis can be realized from its transformation property under the linearized perspective mapping  $F_{*p}$ . With  $B = A_{*q}$  (5) gives

$$\mu_L(q) = G_{*q}^T F_{*p}^T \mu_S(F(G(q))) F_{*p} G_{*q}, \quad (6)$$

where  $p = G(q)$  and  $\mu_S(F(G(q)))$  denotes the second moment matrix defined in the tangent plane to the surface with respect to the window function  $w'(\eta - F(G(q))) = (\det A_{*q})^{-1}w(\xi - q)$ . The general procedure, then, for estimating shape from texture is to combine estimates of  $\mu_L(q)$  with assumptions about the structure of the surface brightness pattern  $\mu_S(F(G(q)))$  in order to infer the structure of  $A_{*q}$ . This permits computation of  $F_{*p}$  after compensation with respect to  $G_{*q}$ .

A simple but often fruitful assumption is that  $\mu_S(F(p))$  is proportional to the unit matrix ( $\tilde{Q} = 0$ ). Such a distribution is called *weakly isotropic*. Under this condition and if  $F_{*p}$  is non-degenerate,  $\mu$  in  $T_p(\Sigma)$  can

be written  $\mu_\Sigma(p) = c F_{*p}^T F_{*p}$  for some  $c > 0$ . Hence, the eigenvectors of  $\mu_\Sigma(p)$  and  $F_{*p}$  are the same, and the eigenvalues of  $F_{*p}$  are proportional to the square roots of the eigenvalues of  $\mu_\Sigma(p)$ . In particular,  $\bar{t} = \pm \bar{e}_1$  and

$$\cos \sigma = \frac{m}{M} = \sqrt{\frac{\lambda_2}{\lambda_1}} = \sqrt{\frac{1 - \tilde{Q}}{1 + \tilde{Q}}}, \quad (7)$$

which gives a direct estimate of local surface orientation if the assumption of weak isotropy can be justified.

A less restrictive assumption is *constant area*. If the local “size” of the surface texture elements does not vary systematically, then  $A = 1/\sqrt{\det \mu}$  is an area measure (since  $\det F_{*p}^{-1} = mM$ ). The normalized *area gradient* defined from it relates to surface orientation by

$$\frac{\nabla A_{\Sigma(p)}}{A_{\Sigma(p)}} = -\tan \sigma \begin{pmatrix} 3 + r\kappa_t / \cos \sigma \\ r\tau \end{pmatrix}, \quad (8)$$

where  $\kappa_t$  is the normal curvature of the surface in the tilt direction, and  $\tau$  is the geodesic torsion in the same direction. From this entity surface orientation can be recovered if the curvature is known or (assumed to be) small. There is no ambiguity in the sign of the tilt direction, unlike the previous case of foreshortening.

### 3 Scale problems in texture analysis

Computation of the image second moment matrix, or any other non-trivial texture descriptor, involves the integration of the image directional statistics over finite-sized local image neighborhoods. This immediately leads to two fundamental scale problems.

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  $s$ ).

Moreover, the image statistics must be based on descriptors computed at proper scales, so that “irrelevant” image structures, e.g. due to noise, can be suppressed. In the method above, which is based on first order spatial derivatives of the image brightness, it is obvious that useful results can hardly be expected if the derivatives are computed directly from unsmoothed noisy data. This scale determining the amount of *initial smoothing* is referred to as *local scale* (denoted  $t$ ).

#### 3.1 Basic idea for scale selection

Given a 2D signal  $f$ , the scale-space representation  $L$  is defined as the solution to the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L = \frac{1}{2} (\partial_{xx} + \partial_{yy}) L \quad (9)$$

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  $g(x, y; t) = (2\pi t)^{-1} \exp(-(x^2 + y^2)/(2t))$ . A well-known property of this representation is that the amplitude of spatial derivatives

$$L_{x^i y^j}(\cdot; t) = \partial_{x^i y^j} L(\cdot; t) = g_{x^i y^j}(\cdot; t) * f(\cdot) \quad (10)$$

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,  $f(x) = \sin \omega_0 x$ , for which the solution to the (1D) diffusion equation is  $L(x; t) = \exp(-\omega_0^2 t/2) \sin \omega_0 x$ . The amplitude  $L_{x^i, max}$  of any  $i$ th order smoothed derivative decreases exponentially with scale

$$L_{x^i, max}(t) = \omega_0^i e^{-\omega_0^2 t/2}. \quad (11)$$

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 is *scale invariance* [10]. In these coordinates the *normalized* (dimensionless) *derivative operator* is

$$\partial_\xi = \sqrt{t} \partial_x. \quad (12)$$

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

$$L_{\xi^i, max}(t) = t^{i/2} \omega_0^i e^{-\omega_0^2 t/2}, \quad (13)$$

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 the 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:

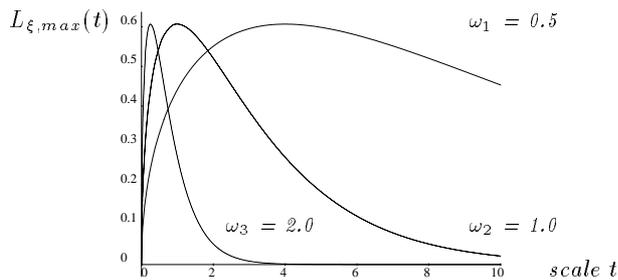
$$\sqrt{t}_{max, L_{\xi^i}} = \frac{\sqrt{i}}{2\pi} \lambda_0. \quad (14)$$

Observe that the maximum value

$$L_{\xi^i, max}(t_{max, L_{\xi^i}}) = i^{i/2} e^{-i/2} \quad (15)$$

is *independent of the frequency* of the signal (see Figure 1). In other words, for these normalized derivatives it holds that sinusoidal signals are treated in a similar (scale invariant) way independent of their frequency.

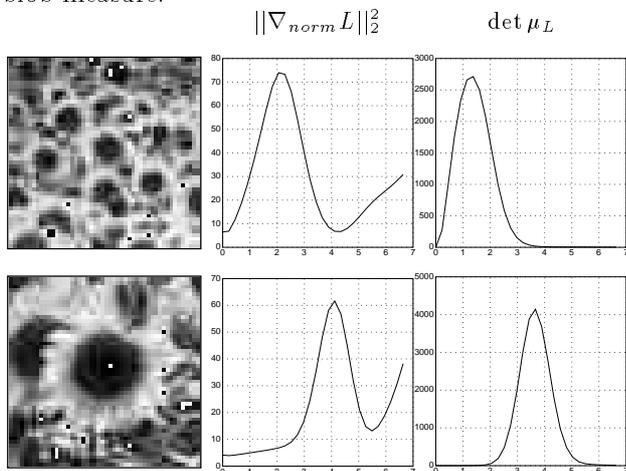
Note the fundamental differences compared to a local Fourier transform; (i) the normalization factor, and (ii) this method allows for local estimates of the frequency content without any explicit setting of window size.



**Figure 1:** 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  $\omega_3 = 2.0$ ).

### 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.* This principle is similar although not equivalent to the parameter variation method in [21, 22], where interesting scale levels are determined from maxima over scales of a (normalized) blob measure.



**Figure 2:** Scale-space signatures of the pointwise normalized gradient magnitude and the determinant of the second moment matrix for two details of a sunflower image; (left) grey-level image, (middle) signature of  $\|\nabla_{norm} L\|_2^2$ , (right) signature of  $\det \mu_L$ . Observe that the maxima in the top row are attained at finer scales than in the bottom row.

Figure 2 illustrates the variation over scale of two 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*,  $\|\nabla_{norm} L\|_2^2$ , and second, the *determinant* of  $\mu_L(q)$ . These graphs are called the *scale-space signatures* of the entities considered.

Clearly, the maxima over scales in the top row of Figure 2 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 [24] for further applications to junction detection and edge detection.

### 3.3 Properties of the 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 [25].

Consider first a sum of two *parallel* (2D) sine waves.  $f_{par}(x, y) = \sin \omega_1 x + \sin \omega_2 x$ , 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,  $f_{orth}(x, y) = \sin \omega_1 x + \sin \omega_2 y$ . 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.

More importantly, *the scale maxima of  $\det \mu_L$  are invariant<sup>1</sup> with respect to linear (and affine) transformations of the image coordinates.* Given an invertible linear transformation  $L(\xi) = R(B\xi)$  it follows from (5)

<sup>1</sup>Here, it is assumed that  $L$  represents an unsmoothed signal and that the scale maximum occurs at zero scale. This restriction is, however, not essential. In [25] it is shown how this result can be extended to strictly positive scales in scale-space.

that  $\det \mu_L = (\det B)^2 \det \mu_R$ , which gives  $\partial_t \det \mu_L = 0 \Leftrightarrow \partial_t \det \mu_R = 0$ .

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  $P_{norm}(\cdot; t) = t(E(L_x^2(\cdot; t)) + E(L_y^2(\cdot; t))) \sim t^{\alpha-1}$ . This expression is independent of scale if and only if  $\alpha = 1$ . In other words, in the 2D case the normalized derivative model is *neutral* with respect to power spectra of the form  $|\omega|^{-2}$ .

Further properties are analysed in [25]. It is shown that when using a Gaussian window function with integration scale  $s$  proportional to the local scale  $t$ , i.e.,  $s = \gamma_1^2 t$ , the sensitivity of the selected scale levels and the average gradient magnitude at that scale is for a periodic sine wave signal proportional to  $\exp(-2\gamma_1^2)$ . The sensitivity is higher for a (Gaussian) blob model.

### 3.4 Uniqueness of the window function

A natural window function to use in (2) is a Gaussian function. It is rotationally symmetric and has a nice scaling behaviour, which means that the invariance properties in Sec. 2.2 are preserved. More importantly, however, then and only then, the components of  $\mu_L$  constitute scale-space representations of the components of  $(\nabla L)(\nabla L)^T$ . This is a direct consequence of the uniqueness of the Gaussian kernel for scale-space representation given natural front-end postulates (e.g., causality [17], or scale invariance [10]). In the rotationally symmetric case, this definition is therefore unique:

$$\mu_L(\cdot; t, s) = g(\cdot; s) * ((\nabla L)(\cdot; t)(\nabla L)(\cdot; t)^T). \quad (16)$$

Of course, separate smoothing of the components of a multi-dimensional entity is not guaranteed to give well-defined (coordinate independent) results. In [25] it is, however, proved that (16) is a meaningful operation.

## 4 Methodology for computing $\mu_L$

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 curve, e.g. the global maximum, or all local extrema, etc;

4. set the integration scale(s) from the scale(s) where the above property is assumed;
5. compute  $\mu_L$  at this 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.

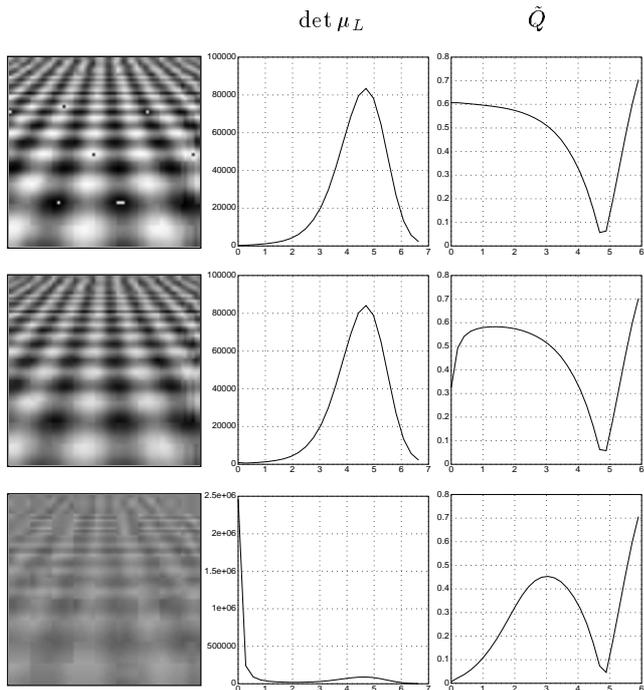
Our specific implementation of this general scheme is described below.

**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.

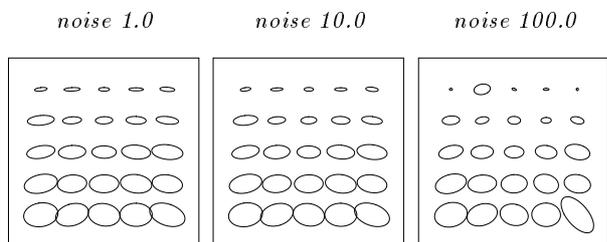
**Selecting integration scales.** Concerning Steps 2–3, we propose to set the *integration scale(s)* from the scale(s), denoted  $s_{\det \mu_L}$ , where the *normalized strength of  $\mu_L$* , represented by  $\det \mu_L$ , assumes a local or global *maximum*. This choice is motivated by the observation 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.

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, this can be motivated by the fact that the estimates of the directional information in  $\mu_L$  are more sensitive to small window sizes than are the magnitude estimates [25].

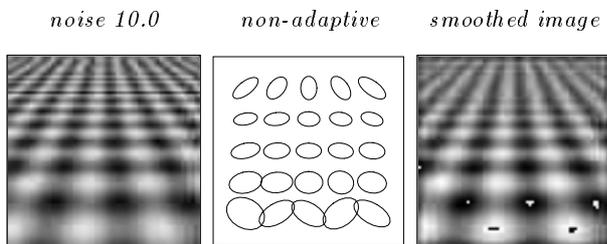
**Selecting local scales.** The second stage selection of local scale in Step 5 aims at reducing the shape distortions due to smoothing. We propose to set the *local scale(s)* to the scale(s), denoted  $t_Q$ , where the *normalized anisotropy*,  $\tilde{Q}$ , assumes a local maximum. This is motivated 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, and it therefore leads to a systematic underestimate of slant; e.g., for the orthographic projection  $f(x, y) = g(x; l_1^2)g(y; l_2^2)$  of a Gaussian blob ( $l_2 = l_1 \cos \sigma$ ), the slant estimate varies as  $\sigma(t) = \arccos((l_2^2 + t)/(l_1^2 + t))^{1/2}$ . On the other hand, suppressing (uniformly distributed) noise and interfering finer scale structures *increases* the anisotropy. Selecting the



**Figure 3:** 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 4:** 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 5:** 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.

maximum point gives a natural trade-off. Note that under the assumption of weak isotropy a maximum in  $\tilde{Q}$  is equivalent to a maximum in  $\sigma$ .

**Experiments.** Fig. 3 illustrates these effects for a synthetic image with different amounts of 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 using the proposed method for scale selection. Notice the stability of  $s_{\det \mu_L}$  with respect to noise.  $t_Q$  increases with  $\nu$ , while  $\tilde{Q}$  decreases at  $t = 0$ . The accuracy in the orientation estimate assuming weak isotropy is presented by the 3D angle  $\Delta\phi_n$  between the estimated and true surface normal. Fig. 4 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 Fig. 5 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.

noise level	$s_{\det \mu_L}$	$t_Q$	$\Delta\phi_n(t_Q)$	$\Delta\phi_n(t=0)$
1.0	34.9	0.0	$0.2^\circ$	$(0.2^\circ)$
10.0	34.4	2.0	$1.1^\circ$	$(15.3^\circ)$
31.6	34.1	4.2	$4.7^\circ$	$(45.3^\circ)$
100.0	31.4	8.5	$7.8^\circ$	$(53.7^\circ)$

**Table 1:** Numerical values of characteristic entities in the experiments in (the center of) Fig. 3 using different amounts of additive Gaussian noise and *automatic* scale selection.

## 5 Spatial selection and blob detection

Although the above method for selecting appropriate scales for smoothing and integration at a given point has been demonstrated to give highly useful results on different types of real and synthetic images [25], it has obvious limitations if applied (blindly) to regions containing little or no image structure. Now, the complementary problem of selecting *where* to apply the multi-scale analysis will be addressed.

The scale selection principle can be easily extended to *spatial selection* by selecting points  $(x, y)$  and scales  $t$  that are local maxima with respect to *both* scale and spatial position. Such points are called *normalized scale-space extrema* of the differential entity considered.

Even though  $\det \mu_L$  can be useful also for spatial selection, it can be desirable to use an entity based on second order derivatives (which typically gives rise to spatial maxima at blob centers) in contrast to first order derivatives (which are zero at extrema). We have

considered the squared trace and the determinant of the normalized Hessian matrix,  $\mathcal{H}_{norm}L$ . For a non-uniform Gaussian blob  $g(x; l_1^2)g(y; l_2^2)$ , the scale maxima are assumed at  $t_{det \mu_L} = l_1 l_2 / (1 + 2\gamma_1^2)^{1/2}$  and  $t_{det \mathcal{H}_{norm}L} = l_1 l_2$ . (The expressions for the corresponding trace-based entities are more complicated although qualitatively similar when  $l_1^2 \approx l_2^2$ ). Observe that  $t_{max}$  gives a natural measure of the area of the blob. In practice, a second stage smoothing step with  $s = \gamma_1^2 t$  is applied to trace  $\mathcal{H}_{norm}L$  and  $\det \mathcal{H}_{norm}L$  before the extrema detection (this is not an essential step, but only a simple way to suppress less significant maxima).

The second column in Fig. 6 shows results of applying this method to real and synthetic images. Every scale-space maximum is illustrated by an ellipse representing  $\mu_L$  computed with integration scale  $s = \gamma_1^2 \gamma_2^2 t_{max}$  and local scale determined from Step 5 in Sec. 4. Note the ability of the method to zoom in to different scales, and how well the computed ellipses describe the blobs in the image, considering how little information is used in the processing. This *multi-scale blob detector* has obvious limitations compared to other approaches [34, 7, 22], 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-from-texture estimation since it produces precisely the information needed for estimating local linear distortion and size changes.

## 6 Estimating surface orientation

The theory developed in the previous sections will now be applied to the problem of estimating local surface orientation in perspective images of textured surfaces. The methodology can be summarized as follows:

**Step 1.** Compute local texture descriptors  $\mu_L$  as described in Sec. 4. This can either be done at selected spatial positions corresponding to normalized scale-space extrema as described in Sec. 5, or at a uniform grid of points generated by some default principle.

**Step 2.** Determine points where surface orientation estimates are to be computed. This set of points can be the same as that used for computing the texture descriptors, or it can be smaller, 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*<sup>2</sup>.

<sup>2</sup>From the semi-group property of Gaussian smoothing it follows that  $\mu_L$  at a coarse scale can be computed from  $\mu_L$  at any finer scale by  $\mu_L(\cdot; t, s_2) = g(\cdot; s_2 - s_1) * \mu_L(\cdot; t, s_1)$ .

**Step 3.** Estimate surface orientation. (a) Apply the assumption of *weak isotropy*, which leads to a direct estimate of surface orientation up to the sign of tilt. (b) Apply the assumption of *constant area*, which permits a unique estimate of surface orientation if the curvature can be neglected in (8). (In practice, the blob area is measured from the scale at which the scale-space maximum is assumed; compare with the analytical results in Sec. 5.) (c) Optionally, compute surface orientation estimates from other texture gradients [12] as well.

**Experiments.** Fig. 6 shows results of applying this composed method to images with known and unknown camera geometry. For the synthetic planar sine wave image (true orientation at center  $(\sigma, \theta) = (60.0^\circ, 90.0^\circ)$ , noise 1.4%) the foreshortening and area gradient estimates are very accurate,  $(60.7^\circ, 90.1^\circ)$  and  $(61.6^\circ, 89.0^\circ)$  respectively. For the synthetic image of the curved cylinder (true orientation  $(55^\circ, 90^\circ)$ , noise 25%) the estimate from foreshortening gives  $(56.8^\circ, 90.4^\circ)$ , while the area gradient underestimates slant  $(36.2^\circ, 86.1^\circ)$ , since the curvature is non-zero (compare with (8)). For the real (planar) wall-paper image, foreshortening gives  $(47.9^\circ, 84.6^\circ)$  and the area gradient  $(51.4^\circ, 76.6^\circ)$  to be compared to the reference value  $(50.8^\circ, 85.3^\circ)$  computed by stereo correspondence.

Finally, for the sunflower image [7] the camera geometry is unknown, so the area gradient is represented by a set of lines determining the horizon (which is independent of the focal length). Note that for this image the weak isotropy assumption gives the orientation of the individual flowers, while the constant area assumption reflects the orientation of the underlying surface.

## 7 Summary and discussion

This paper addresses the shape-from-texture problem from a general standpoint. Results from previous and new theories are combined and extended into a framework that, in fact, models all the steps in the computation of local slant and tilt information from raw grey-level data. The method is general in the sense that the description stage is decoupled from any specific assumptions about the type of texture considered; in fact, all such choices are postponed to the interpretation stage, where different assumptions can be selected depending upon the type of situation.

Concerning limitations of the work, the genuine two-parameter scale variation has so far not been implemented in the selection of integration scale. The current implementation gives best results when either the textures are approximately periodic, the relative integration scale is known (given by  $\gamma_1$  and  $\gamma_2$ ), or the image contains relatively distinct blob-like structures.

**Multiple scale maxima and coherent surfaces.** In the experiments only the most dominant scale was selected at each point, but it is clearly desirable to be able to handle patterns containing structures at multiple scales. Moreover, the experiments primarily concerned local (or regional) estimation of surface orientation; some mechanism is needed for combining such (agreeing and conflicting) slant and tilt estimates computed at different points into hypotheses about coherent surfaces.

**Non-uniform smoothing.** In the treatment above *uniform* (rotationally symmetric) smoothing has been used throughout. This is motivated by the principle that in the absence of any a priori information, the vision system should be as *uncommitted* as possible, and, for example, have no preferred directions. It is clear, however, that such uniform smoothing leads to shape distortions (compare with Sec. 4; see also [32]).

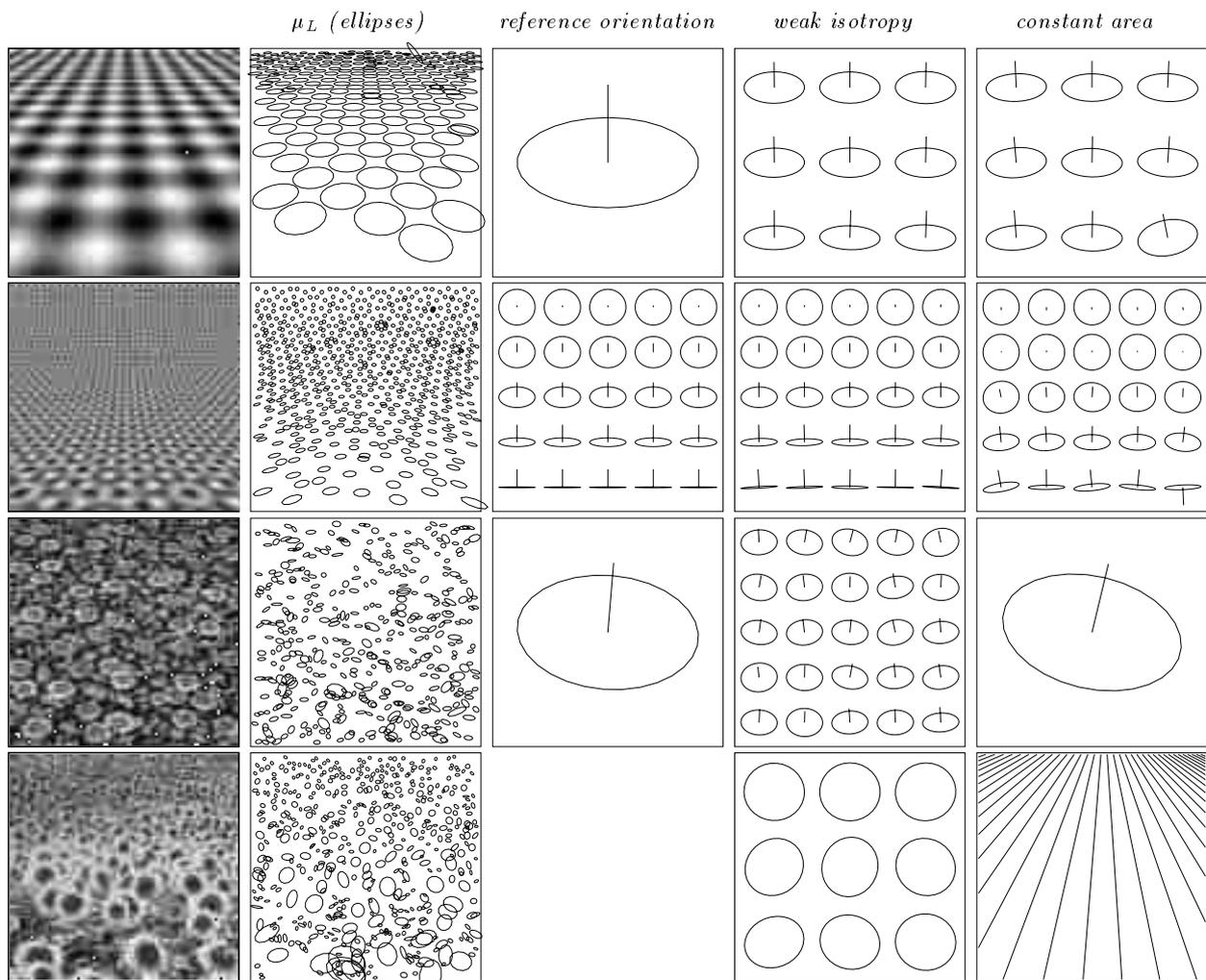
Given initial slant and tilt estimates ( $\hat{\sigma}, \hat{\theta}$ ) computed by uniform smoothing, a straightforward compensation technique is to let the scale parameters in the (estimated) tilt direction,  $t_{\hat{t}}$  and  $s_{\hat{t}}$ , and the perpendicular direction,  $t_{\hat{b}}$  and  $s_{\hat{b}}$ , be related by  $t_{\hat{t}} = t_{\hat{b}} \cos^2 \hat{\sigma}$  and  $s_{\hat{t}} = s_{\hat{b}} \cos^2 \hat{\sigma}$ . If, say, under the assumption of weak isotropy, this estimate is correct, then the slant estimate will be *unaffected* by this non-uniform smoothing. Applying (one iteration) of this method to the experiment in Table 1 gives that at noise levels 10.0, 31.6 and 100.0, the slant estimates change from  $58.9^\circ$  to  $59.9^\circ$ , from  $55.3^\circ$  to  $61.1^\circ$ , and from  $52.3^\circ$  to  $62.9^\circ$  respectively (the reference value is  $60^\circ$ ). It is not argued that this method describes any “optimal” way to reduce the shape distortions. The intention is rather to demonstrate that *shape adaption* of the smoothing kernels can be performed within this framework. A formal description in [25] shows that the entire shape-from-texture method becomes *invariant with respect to the locally linearized perspective mapping*. Incidentally, this provides an alternative way to define anisotropic diffusion [30].

**Active vision.** The effect of the smoothing operation is closely related to the issue of view direction. Ideally, in noise free and unsmoothed data it should be sufficient to compensate  $\mu_L$  computed at a peripheral point  $q$  using (6) and the (known) gaze transformation  $G_{*q}$ . However, the isotropic smoothing in the flat image and the use of non-infinitesimal integration domains lead to non-linear effects not covered by this. A natural solution is to map the intensity to the viewsphere, and perform the diffusion there in order to obtain invariance with respect to view direction. In an active vision situation, however, the situation can also be improved by redirecting the camera towards the point of interest, since then the peripheral perspective effects are as smallest,

and symmetric around the point considered. Note that under the weak isotropy assumption the slant and tilt estimates at the central point are independent of the focal length.

## References

- [1] J. Aloimonos, “Shape from texture”, *Biol. Cyb.*, 58, 345–360, 1988.
- [2] J. Babaud, A. P. Witkin, M. Baudin, R. O. Duda, “Uniqueness of the Gaussian kernel for scale-space filtering”, *IEEE-TPAMI*, 8:1, 26–33, 1986.
- [3] J.R. Bergen, E.H. Adelson, “Early vision and texture perception”, *Nature*, 333, 363–364, 1988.
- [4] J. Bigün, G.H. Granlund, J. Wiklund, “Multidimensional orientation estimation with applications to texture analysis and optical flow”, *IEEE-TPAMI*, 13, 775–790, 1991.
- [5] A. Blake, C. Marinos, “Shape from texture: estimation, isotropy and moments”, *J. Art. Intell.*, 45, 323–380, 1990.
- [6] A. Blake, C. Marinos, “Shape from texture: the homogeneity hypothesis”, *3rd ICCV*, Osaka, Japan, 350–353, 1990.
- [7] D. Blostein, N. Ahuja, “Shape from texture: integrating texture element extraction and surface estimation”, *IEEE-TPAMI*, 11, 1233–1251, 1989.
- [8] L. G. Brown, H. Shvaytser, “Surface orientation from projective foreshortening of isotropic texture autocorrelation”, *IEEE-TPAMI*, 12, 584–588, 1990.
- [9] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, M. A. Viergever, “General intensity transformations”, *7th SCIA*, (Aalborg, Denmark), 338–345, 1991.
- [10] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, M. A. Viergever, “Scale and the differential structure of images”, *Image Vision Comp.*, 10:6, 376–388, 1991
- [11] M.A. Förstner, E. Gülch, “A fast operator for detection and precise location of distinct points, corners and centers of circular features”, *ISPRS Intercomm. Workshop*, 1987.
- [12] J. Gårding, “Shape from texture for smooth curved surfaces in perspective projection”, *J. Math. Imaging Vision*, 2, 329–352, 1992.
- [13] J. Gårding, “Shape from texture and contour by weak isotropy”, *J. Art. Intell.*, 1993. (In press).
- [14] J. Gibson, *The Perception of the Visual World*. Houghton Mifflin, Boston, 1950.
- [15] K. Kanatani, T.C. Chou, “Shape from texture: general principle”, *J. Art. Intell.*, 38, 1–48, 1989.
- [16] H. Knutsson, G.H. Granlund, “Texture analysis using two-dimensional quadrature filters”, *IEEE Workshop on Comp. Arch. Patt. Anal. Image Database Management*, 1983.
- [17] J.J. Koenderink, “The structure of images”, *Biol. Cyb.*, 50, 363–370, 1984.
- [18] J. J. Koenderink, A. J. van Doorn, “Receptive field families”, *Biol. Cyb.*, 63, 291–298, 1990.
- [19] A.F. Korn, “Toward a symbolic representation of intensity changes in images”, *IEEE-TPAMI*, 10:5, 610–625, 1988.
- [20] T. Lindeberg, “Scale-space for discrete signals”, *IEEE-TPAMI*, 12, 234–254, Mar. 1990.
- [21] T. Lindeberg, *Discrete scale space theory and the scale space primal sketch*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 1991. A revised version to appear in Kluwer Int. Ser. in Engineering and Computer Science.
- [22] T. Lindeberg, J.-O. Eklundh, “Scale-space primal sketch: Computation and Experiments”, *Image Vision Comp.*, 10, 3–18, 1992.



**Figure 6:** Local surface orientation estimates computed by the composed shape-from-texture method; (row 1) synthetic image of a planar surface with 1.4% noise, (row 2) synthetic image of a curved surface with 25% noise, (row 3) real image of a planar surface with known orientation, (row 4) real image of a sunflower field with unknown orientation. The columns show from left to right; (a) grey-level image, (b) blobs detected by the multi-scale blob detection method, (c) reference surface orientation (if known), (d) surface orientation assuming weak isotropy, (e) surface orientation assuming constant area.

- [23] T. Lindeberg, "Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction", *J. Math. Imaging Vision*, 1993. (To appear).
- [24] T. Lindeberg, "On scale selection for differential operators". *8th SCIA*, (Tromsø, Norway), 1993. (To appear).
- [25] T. Lindeberg, J. Gårding, "Shape from texture from a multi-scale perspective", *Technical Report*, Royal Institute of Technology, S-100 44 Stockholm, Sweden, Jan 1993.
- [26] J. Malik, P. Perona, "A computational model of texture segmentation", in *CVPR*, (San Diego, Ca.), 1989.
- [27] K.V. Mardia, *Statistics of Directional Data*. Academic Press, (London), 1972.
- [28] D. Marr, "Early processing of visual information", *Phil. Trans. Royal Soc (B)*, 275, 483-524, 1976.
- [29] A.P. Pentland, "Shading into texture", *J. Art. Intell.*, 29, 147-170, 1986.
- [30] P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE-TPAMI*, 12:7, 629-639, 1990.
- [31] A. R. Rao, B. G. Schunk, "Computing oriented texture fields", *CVGIP*, 53, 157-185, 1991.
- [32] J.V. Stone, "Shape from texture: textural invariance and the problem of scale in perspective images of surfaces", *BMVC*, (Oxford, England), 1990.
- [33] B.J. Super, A.C. Bovik, "Shape-from-texture by wavelet-based measurement of local spectral moments", *CVPR*, (Urbana-Champaign, Illinois), 296-301, 1992.
- [34] H. Voorhees, T. Poggio, "Detecting textures and texture boundaries in natural images", *1st ICCV*, (London), 1987.
- [35] A.P. Witkin, "Recovering surface shape and orientation from texture", *J. Art. Intell.*, 17, 17-45, 1981.
- [36] A.P. Witkin, "Scale-space filtering", *8th IJCAI*, (Karlsruhe, Germany), 1019-1022, 1983.
- [37] A. L. Yuille, T. A. Poggio, "Scaling theorems for zero-crossings", *IEEE-TPAMI*, 8, 15-25, 1986.
- [38] W. Zhang, F. Bergholm, "An extension of Marr's "signature" based edge classification", *7th SCIA*, (Aalborg, Denmark), 435-443, 1991.