# Shape-adapted smoothing in estimation of 3-D shape cues from affine distortions of local 2-D brightness structure*

Tony Lindeberg and Jonas Gårding

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@@nada.kth.se, jonasg@@bion.kth.se

(Original version Oct. 1993; Revised Jan. 1994, Sep. 1994, Oct. 1996)

March 13, 2001

**Abstract**

This article describes a method for reducing the shape distortions due to scale-space smoothing that arise in the computation of 3-D shape cues using operators (derivatives) defined from scale-space representation. More precisely, we are concerned with a general class of methods for deriving 3-D shape cues from 2-D image data based on the estimation of locally linearized deformations of brightness patterns. This class constitutes a common framework for describing several problems in computer vision (such as shape-from-texture, shape-from-disparity-gradients, and motion estimation) and for expressing different algorithms in terms of similar types of visual front-end-operations.

It is explained how surface orientation estimates will be biased due to the use of rotationally symmetric smoothing in the image domain. These effects can be reduced by extending the linear scale-space concept into an *affine Gaussian scale-space* representation and by performing *affine shape adaptation* of the smoothing kernels. This improves the accuracy of the surface orientation estimates, since the image descriptors, on which the methods are based, will be relative invariant under affine transformations, and the error thus confined to the higher-order terms in the locally linearized perspective transformation.

A straightforward algorithm is presented for performing shape adaptation in practice. Experiments on real and synthetic images with known orientation demonstrate that in the presence of moderately high noise levels the accuracy is improved by typically one order of magnitude.

*A short version of this article has been published in Proc. 3rd European Conference on Computer Vision, (Stockholm, Sweden), May 1994, pp. 389–400, vol. 800 of Springer-Verlag Lecture Notes in Computer Science.

# 1 Introduction

This article deals with the problem of estimating cues to local three-dimensional scene structure using textural properties measured from either a single monocular image or a binocular image pair. The notion of *scale* is of crucial importance in this context, and we consider the problem of adapting traditional scale-space theory to the shape estimation process, thereby reducing bias effects due to rotationally symmetric smoothing.

## 1.1 Scale-space representation

To derive any information from image data it is necessary to interact with it using operators. Some of the very fundamental questions in computer vision concern what operators to use, how large they should be, and where they should be applied. If these problems are not properly dealt with, the task of interpreting the operator response can be very hard.

A systematic approach that has been developed to address the problem of what operators to use is *scale-space theory*. It focuses on the basic property of image data that image structures, in general, exist at different scales and the fact that one cannot expect to know in advance at what scales relevant image structures exist. A fundamental assumption is that in cases when no *a priori* information is available, the only reasonable approach is to treat image structures at all scales simultaneously and as uniformly as possible. Analogously, all image points should be treated in a similar manner.

Starting from these basic properties several axiomatic derivations have been given concerning what image operators to use (Witkin 1983, Koenderink 1984, Babaud, Witkin, Baudin & Duda 1986, Yuille & Poggio 1986, Lindeberg 1990, Lindeberg 1993*a*, Lindeberg 1994*c*, Lindeberg 1994*a*, Koenderink & van Doorn 1990, Florack, ter Haar Romeny, Koenderink & Viergever 1992). The essence of these results is that if one assumes that the first stages of visual processing should be as *uncommitted* as possible and have no particular bias, then, within the class of linear transformations, convolution with Gaussian kernels and their derivatives is singled out as a canonical class of low-level operations. The output from these operators can in turn be used as a basis for a large number of early visual operations, such as feature detection, matching, and computation of shape cues.

It is, however, well-known that shape distortions may arise in image descriptors defined from the scale-space representation based on the rotationally symmetric Gaussian kernel and its derivatives. In edge detection, for example, smoothing across "object boundaries" can affect both the shape and the localization of edges. A corresponding problem arises in the derivation of three-dimensional shape cues from image data. In shape-from-texture, for example, rotationally symmetric smoothing affects the local anisotropy in the image (measured, for instance, in terms of the distribution of gradient directions), which means that surface orientation estimates may be biased. A common effect that occurs in practice is that the slant angle (the angle between the visual ray and the surface normal) is systematically underestimated.

## 1.2 Non-uniform smoothing methods

To reduce the problems of shape distortion in edge detection, (Perona & Malik 1990) proposed the use of *anisotropic diffusion* as a generalization of the linear scale-space representation (which is generated by the (linear) diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L = \frac{1}{2} \sum_{i=1}^{N} \partial_{x_i x_i} L \tag{1}$$

with initial condition $L(\cdot;\ 0) = f$, where $f$ denotes the original signal). The basic idea is to modify the conductivity $c(x;\ t)$ in a non-linear version of the diffusion equation $\partial_t L = \frac{1}{2} \nabla^T (c(x;\ t) \nabla L)$ such as to favour intra-region smoothing to inter-region smoothing. In principle, they solved the diffusion equation

$$\partial_t L = \tfrac{1}{2} \nabla^T (h(|\nabla L(x;\ t)|) \nabla L) \tag{2}$$

for some monotonic decreasing function $h\colon \mathbb{R}_+ \to \mathbb{R}_+$. The intuitive effect of this evolution is that the conductivity will be low where the gradient magnitude is high and vice versa.

This idea has been further developed by several authors. (Nordström 1990) showed that by adding a bias term to the diffusion equation, it was possible to relate this method to earlier considered regularization approaches by (Terzopoulos 1983) and (Mumford & Shah 1985). Alternative modifications in terms of adaptive smoothing schemes have been presented by (Saint-Marc, Chen & Medioni 1991, Nitzberg & Shiota 1992).

By adopting an axiomatic approach, (Alvarez, Guichard, Lions & Morel 1993) have shown that given certain constraints on a visual front-end, a natural choice of non-linear diffusion equation is the equation

$$\partial_t L = \tfrac{1}{2} |\nabla L| \nabla^T (\nabla L / |\nabla L|) = \tfrac{1}{2} L_{\bar{u}\bar{u}}, \tag{3}$$

where $L_{\bar{u}\bar{u}}$ represents the second order derivative in the tangent direction to a level curve. This evolution, also used by (Osher & Sethian 1988) and (Kimia, Tannenbaum & Zucker 1990), means that level curves move in the normal direction with a velocity proportional to the curvature of the level curves. For a slightly modified version of (3),

$$\partial_t L = \tfrac{1}{2} (|\nabla L|^2 L_{\bar{u}\bar{u}})^{1/3}, \tag{4}$$

the solutions are relative invariant under affine transformations of the spatial coordinates (Alvarez et al. 1993). This property has been used by (Sapiro & Tannenbaum 1993) for defining an affine invariant curve evolution scheme.

An interesting approach to describing non-linear diffusion more generally is pursued by (Florack, Salden, t. Haar Romeny, Koenderink & Viergever 1995), who consider general non-linear coordinate transformations of the spatial coordinates as a common framework for expressing such operations. Interestingly, this approach covers several of the above-mentioned methods. A more general overview of non-uniform smoothing approaches can be found in the book edited by (ter Haar Romeny 1994).

## 1.3 Affine shape-adapted smoothing

Improvements relative to the rotationally symmetric scale-space representation can also be obtained using linear theory. As has been argued by several authors, it can

in certain situations be advantageous to use filters that correspond to different scale values along different directions; for example a large scale value along the direction of an edge, and a smaller scale value in the perpendicular direction. At junctions, on the other hand, where several directions meet, a converse behaviour may be desired to better resolve the directional information.

In this article we shall develop how linear, or affine, shape adaptation of the smoothing kernels can be used as an important mechanism for reducing the shape distortions that occur when using the traditional rotationally symmetric scale-space representation for computing three-dimensional shape cues from image data. The basic underlying approach we shall adopt for relating image data to local surface shape is to observe how surface patterns are distorted under projective transformations. This problem can be substantially simplified by approximating the projective transformation with its locally linearized transformation (the derivative).

The basic idea of using measurements of local linear, or affine, distortions of two-dimensional patterns as a cue to three-dimensional structure goes back to (at least) (Koenderink & van Doorn 1976), and has been applied to several shape-from-X problems, such as shape-from-texture (Gårding 1992, Lindeberg & Gårding 1993, Malik & Rosenholtz 1993), shape-from-disparity (Wildes 1981, Jones & Malik 1992, Gårding & Lindeberg 1996), and motion estimation (Koenderink & van Doorn 1991, Bergen, Anandan, Hanna & Hingorani 1992, Weber & Malik 1993, Cipolla, Okamoto & Kuno 1993). Related works on the estimation of affine transformations have been presented by (Manmatha 1994, Sato & Cipolla 1994).

The advantages of shape-adapted smoothing in shape-from-texture were first pointed out by (Stone 1990), who proposed to adapt the smoothing kernels to be isotropic when backprojected to the surface, rather than in the image. He also suggested an iterative scheme, based on the principle of first estimating the surface orientation and then adapting the kernel shape accordingly.

Here, we shall apply and generalize this idea, and demonstrate how the accuracy of shape-from-X methods based on local affine distortion can be improved by extending the rotationally symmetric *linear scale-space* representation to an *affine Gaussian scale-space* representation based on Gaussian kernels with arbitrary (positive definite) covariance matrices. The methods we shall consider here are based on a specific image descriptor called the *second moment matrix*, which reflects the local distribution of gradient directions in the image. However, the need for shape adaptation arises in many other contexts (such as stereo matching over windows) if one aims at image descriptors that obey transformation properties exactly and do not bias measurements between corresponding image patches.

## 2 Intuitive idea for shape adaptation

To motivate the need for shape adaptation and to give an intuitive illustration of the basic idea behind the presented approach, let us study an idealized model pattern for the effect of scale-space smoothing can be analysed in closed form. Consider first a non-uniform Gaussian blob

$$f(x, y) = g(x;\ l_1^2)\, g(y;\ l_2^2) \quad (l_1 \geq l_2 > 0), \tag{5}$$

as a simple linearized model of the projection of a rotationally symmetric Gaussian blob. The foreshortening, $\epsilon$, and the slant angle, $\sigma$, are given by

$$\epsilon = \cos\sigma = \frac{l_2}{l_1}, \tag{6}$$

and the tilt direction (the direction of the projection of the surface normal onto the image plane) is $\theta = \pi/2$. From the semi-group property of the Gaussian kernel $g(\cdot;\ t_1) * g(\cdot;\ t_2) = g(\cdot;\ t_1 + t_2)$, it follows that the scale-space representation of $f$ at scale $t$ is

$$L(x, y;\ t) = g(x;\ l_1^2 + t)\, g(y;\ l_2^2 + t). \tag{7}$$

This means that under scale-space smoothing foreshortening varies as

$$\epsilon(t) = \sqrt{\frac{l_2^2 + t}{l_1^2 + t}}, \tag{8}$$

i.e., it increases and tends to one, which means that after a sufficiently large amount of smoothing the image will eventually be interpreted as flat.

On the other hand, if we have initial estimates of the slant angle and the tilt direction $(\hat{\sigma}, \hat{\theta})$, say computed using rotationally symmetric Gaussian smoothing, a straightforward compensation technique is to let the scale parameter in the (estimated) tilt direction, denoted $t_{\hat{t}}$ and the scale parameter in the perpendicular direction, denoted $t_{\hat{b}}$ be related by

$$t_{\hat{t}} = t_{\hat{b}}\, \cos^2\hat{\sigma}. \tag{9}$$

If this estimate is correct, then the slant estimate computed from the image data in the shape-adapted scale-space representation will be *unaffected* by the non-uniform smoothing operation. To algebraically illustrate this property, assume first, for simplicity, that the estimate of the tilt direction is correct ($\hat{\theta} = \theta = \pi/2$) and convolve the signal with a non-uniform Gaussian kernel

$$g(x, y;\ t_{\hat{t}}, t_{\hat{b}}) = g(x;\ t_{\hat{t}})\, g(y;\ t_{\hat{b}}). \tag{10}$$

From the semi-group property of the Gaussian kernel it follows that the shape-adapted scale-space representation of $f$ using non-uniform smoothing is

$$L(x, y;\ t) = g(x;\ l_1^2 + t_{\hat{b}})\, g(y;\ l_2^2 + t_{\hat{t}}), \tag{11}$$

and the new foreshortening estimate corresponding to (8) is

$$\hat{\epsilon} = \epsilon(\hat{\sigma};\ t_{\hat{t}}, t_{\hat{b}}) = \sqrt{\frac{l_2^2 + t_{\hat{t}}}{l_1^2 + t_{\hat{b}}}} = |\cos\sigma| \sqrt{1 + \frac{t_{\hat{b}}}{l_1^2 + t_{\hat{b}}} \left( \frac{\cos^2\hat{\sigma}}{\cos^2\sigma} - 1 \right)}. \tag{12}$$

It is clear that if the initial slant estimate is correct ($\hat{\sigma} = \sigma$), then the estimate of foreshortening is correct ($\hat{\epsilon} = \epsilon$). This means that surface orientation estimate will be unaffected by the smoothing operation (and corresponds to rotationally symmetric smoothing in the tangent plane to the surface).

In practice, one cannot, of course, assume that correct values of $(\sigma, \theta)$ are known, since this requires knowledge about the solution to the problem we are to solve. A more realistic formulation of the problem it therefore to first compute initial surface

orientation estimates using rotationally symmetric smoothing (based on the principle that in situations when no *a priori* information is available, the first stages of visual processes should be as uncommitted as possible and have no particular bias). (For this specific example the tilt angle will be correct, whereas the slant angle will be underestimated if a non-zero value of $t$ is used). Then, when a hypothesis about a certain surface orientation $(\hat{\sigma}_0, \hat{\theta}_0)$ has been established, an iterative method for shape adaptation can be formulated based on successive estimates of the surface normal,

$$\hat{\sigma}_{n+1} = \arccos \epsilon(\hat{\sigma}_n;\ t_{\hat{t}}, t_{\hat{b}}) = h(\hat{\sigma}_n). \tag{13}$$

From the derivative of this mapping,

$$|(\partial_{\hat{\sigma}} h)(\hat{\sigma})| = |(\partial_{\hat{\sigma}} \arccos \epsilon)(\hat{\sigma};\ t_{\hat{b}} \cos \hat{\sigma}, t_{\hat{b}})| = \{\text{let } \hat{\sigma} = \sigma\} = \frac{t_b}{l_1^2 + t_b} < 1,$$

it is clear that the true value of $\hat{\sigma}$ is a *convergent fixed point* of the above equation (with $\epsilon = \cos \hat{\sigma}$). Hence, in this example, the method is guaranteed to converge to the true solution, provided that the initial estimate is sufficiently close to the true value.

Although this example concerns a specific example (orthographic projection of a rotationally symmetric surface pattern) it illustrates the following general properties:

- Isotropic smoothing in the image domain distorts the two-dimensional "shape" of the image pattern. Unless this effect is taken into account, *i.e.* unless shape adaptation is performed, the surface orientation estimates will be biased.

  This problem can be expected to arise for any shape-from-X method based on operators defined from the rotationally symmetric Gaussian kernel.

- A fixed-point property is a natural indicator of successful shape adaptation.

  More generally, a fixed-point property reflects the fact that the shape of the smoothing kernel agrees with the shape of the image structure, and is therefore independent any specific assumption about the surface texture (*e.g.* isotropy). If such a fixed-point is preserved under linear transformations, then image descriptors computed at the fixed point have a predictable behaviour under linear transformations, and it is possible to preserve transformation properties exactly.

In the following, we shall develop these ideas further and apply them to the problems of estimating local surface orientation from either monocular texture distortion or from the gradient of binocular disparity. The specific shape-from-X methods we shall be concerned with are based on an image descriptor called *the windowed second moment matrix*, which is reviewed in section 3. The *affine Gaussian scale-space* representation described in section 4 provides the necessary theoretical basis for defining a shape-adapted second moment descriptor in section 5. In section 6, an iterative procedure for computing this descriptor is proposed, and in section 7 it is reviewed how shape-from-texture and shape-from-disparity-gradients algorithms can be expressed in terms of second moment matrices. Experimental results of including shape adaptation in these schemes are presented in section 8, and section 9 summarizes the approach.

## 3 The windowed second moment descriptor

Several local shape estimation problems can be formulated in terms of estimating the parameters of a local affine[1] transformation, either between the image and a surface patch with known structure, or between two image patches. The second moment matrix described in this section is a highly useful descriptor for this purpose; in principle it makes it possible to express the transformation parameters directly in terms of the parameters of the computed descriptors. The second moment matrix also fits naturally within the scale-space framework.

Because the second moment descriptor has only three degrees of freedom, it is not possible to estimate all parameters of a general linear transformation; the transformation is determined up to an arbitrary rotation. (More precisely, given two positive definite second moment matrices $\mu_L$ and $\mu_R$, the transformation property (17) gives

$$B = \mu_R^{-1/2} R \, \mu_L^{1/2}, \tag{14}$$

where $\mu^{1/2}$ represents any matrix that satisfies $(\mu^{1/2})^T(\mu^{1/2}) = \mu$ and $R$ is an arbitrary rotation matrix.) Fortunately, however, further geometric constraints can in many cases be exploited to eliminate this ambiguity.

In this section, we shall review the definition of the basic (non-adapted) second moment descriptor based on isotropic smoothing. After introducing the affine Gaussian scale-space representation in section 4, we shall then in section 5 define a second moment descriptor that allows for shape adaptation.

### 3.1 Measuring local affine distortion

Let $L\colon \mathbb{R}^2 \to \mathbb{R}$ denote the image brightness and let $\nabla L = (L_x, L_y)^T$ be its gradient. Given a symmetric and normalized window function $w$, *the windowed second moment matrix* $\mu\colon \mathbb{R}^2 \to \mathrm{SPSD}(2)$[2] can be defined as

$$\begin{aligned} \mu_L(q) &= \int_{x \in \mathbb{R}^2} (\nabla L(x))(\nabla L(x))^T \, w(q - x) \, dx \\ &= \int_{(x_1, x_2) \in \mathbb{R}^2} \begin{pmatrix} L_{x_1}^2 & L_{x_1} L_{x_2} \\ L_{x_1} L_{x_2} & L_{x_2}^2 \end{pmatrix} w(q_1 - x_1, q_2 - x_2) \, dx_1 \, dx_2, \end{aligned} \tag{15}$$

where $q = (q_1, q_2) \in \mathbb{R}^2$ denotes the image point at which it is computed. Introduce an averaging operator $E_q$ describing the effect of integration with the window function centered at $q$. Then, (15) can more compactly be written

$$\mu_L(q) = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} = E_q \begin{pmatrix} L_x^2 & L_x L_y \\ L_x L_y & L_y^2 \end{pmatrix} = E_q((\nabla L)(\nabla L)^T). \tag{16}$$

This type of descriptor can be thought of as the covariance matrix of a two-dimensional stochastic variable, or as the moment of inertia of a mass distribution in the plane. Different versions of it have been used by several authors, for example, (Förstner & Gülch 1987, Bigün 1990, Brown & Shvaytser 1990, Blake & Marinos 1990*a*, Bigün,

[1] For the problems considered in this article, the translational part of the affine transformation is either irrelevant or will be assumed to be known. Therefore, the terms "affine" and "linear" are used interchangeably.

[2] The notation SPSD(2) stands for the cone of symmetric positive semidefinite $2 \times 2$ matrices.

Granlund & Wiklund 1991, Rao & Schunk 1991, Super & Bovik 1992, Gårding 1992, Gårding 1993, Lindeberg & Gårding 1993).

The usefulness of this descriptor with respect to estimation of linear distortion can be realized from its transformation property under linear transformations. Let $B$ be an invertible linear transformation of the image domain (represented by an invertible $2 \times 2$ matrix), and define a transformed intensity pattern $R\colon \mathbb{R}^2 \to \mathbb{R}$ by $L(\xi) = R(B\xi)$. Then, it can be shown that $\mu_L(q)$ transforms according to

$$\mu_L(q) = B^T \, \mu_R(p) \, B, \tag{17}$$

where $\mu_R(p)$ is the second moment matrix of $R$ at $p = Bq$ computed with respect to the "backprojected" normalized window function

$$w'(\eta - p) = (\det B)^{-1} w(\xi - q) \tag{18}$$

and $\eta = B\xi$. This result can be easily verified by inserting $\nabla L(\xi) = B^T \nabla R(B\xi)$ into (16), which gives

$$\begin{aligned}\mu_L(q) =& \int_{\xi \in \mathbb{R}^2} w(q - \xi) \, B^T (\nabla R(B\xi))(\nabla R(B\xi))^T B \, d\xi \\ =& B^T \left\{ \int_{\eta \in \mathbb{R}^2} w(B^{-1}(p - \eta)) \, (\nabla R(\eta))(\nabla R(\eta))^T \, (\det B)^{-1} \, d\eta \right\} B.\end{aligned}$$

The integral within brackets can be recognized as the second moment matrix of $R$ at $p$ if $w'$ according to (18) is used as window function. Clearly, the transformed window function $w'$ is normalized if $w$ is, since

$$\int_{\eta \in \mathbb{R}^2} w(B^{-1}(\eta - p))(\det B)^{-1} \, d\eta = \int_{\xi \in \mathbb{R}^2} w(\xi - q) \, d\xi.$$

### 3.2 Representing and selecting scale

Computation of the second moment matrix from image data (or any other non-trivial texture descriptor) involves the integration of local image statistics (here, gradient directions) over finite-sized image regions. This immediately leads to two scale problems. The first problem concerns the scale(s) at which to compute the primitives for the texture descriptors (here, the first order derivatives). This scale determining the amount of *initial smoothing* in the (traditional first-stage) scale-space representation of the image is called *local scale* (denoted $t$). The second scale problem concerns the size of the regions over which to collect the statistics. This scale controlling the size of the *window function* (which is selected as a Gaussian function) is referred to as *integration scale* (denoted $s$).

With respect to these two scale concepts, the brightness function $L$ in (15) can be reinterpreted as the scale-space representation of any given image $f$, and the window function $w$ be associated with a scale parameter. This gives the *multi-scale windowed second moment matrix*

$$\mu_L(q;\ t, s) = \int_{x' \in \mathbb{R}^2} (\nabla L)(x';\ t) \, (\nabla L)^T(x';\ t) \, g(q - x';\ s) \, dx'. \tag{19}$$

This descriptor depends on two scale parameters. To use it for distortion measurements it is necessary to have some mechanism for selecting scale levels automatically,

since scales useful for measuring affine transformations cannot, in general, be assumed to be known in advance. In typical shape-from-X problems, such scale levels can be expected to vary substantially over the image, depending on the type of texture considered, the distance to the surface, and the noise in the image formation process.

One general principle for automatic scale selection has been proposed in (Lindeberg 1993*b*, Lindeberg 1994*b*). The basic idea is to study the evolution over scales of (possibly non-linear) combinations of normalized derivatives defined by

$$\partial_{\xi_i} = \sqrt{t}\, \partial_{x_i}, \tag{20}$$

(where $\xi = x/\sqrt{t}$ are normalized coordinates) and to select scale levels from the scales at which such entities based on normalized derivatives assume local maxima over scales. In junction detection and edge detection, useful entities for selecting detection scales are the normalized rescaled curvature of level curves, $\tilde{\kappa}_{norm} = t^2\, |\nabla L|^2\, L_{\bar{u}\bar{u}}$, and the normalized gradient magnitude, $L_{v,\gamma-norm} = \sqrt{t}\, L_v$, respectively.

With respect to the problem of computing the second moment matrix descriptors, a useful strategy is to *select integration scales* at a given point from the scale at which the determinant or the trace of the normalized second moment matrix (Lindeberg & Gårding 1993, Gårding & Lindeberg 1996)

$$\det \mu_{L,norm} = t^2\, \det \mu_L \quad \text{trace}\, \mu_{L,norm} = t\, \text{trace}\, \mu_L \tag{21}$$

assumes local maxima over scales. (In practice, the scale is selected as some constant $\gamma \geq 1$ times this scale value, typically $\gamma = 1$, $\sqrt{2}$, or 2). Ideally, this means that the size of the image patch for which the descriptor is computed reflects the characteristic size of the significant image structures contained in the patch.

Then, given a value of the integration scale, a value for the *local scale parameter* must be selected. If the noise level is known and the perspective effects are small, it may be sufficient to use a fixed value. A more sophisticated approach is to select the local scale(s) for which the normalized anisotropy

$$\tilde{Q} = \frac{\sqrt{\text{trace}^2\, \mu_L - 4 \det \mu_L}}{\text{trace}\, \mu_L}, \tag{22}$$

assume local maxima over scales. The motivation for this choice is that suppression of (isotropic) noise is likely to lead to increasing anisotropy, whereas shape distortions due to isotropic smoothing are likely to lead to decreasing anisotropy. Selecting a local maximum over scales gives a natural balance between these two effects.

An alternative methodology suitable for blob-like textures (Lindeberg 1993*b*, Lindeberg 1994*b*)[3] is to simultaneously select interesting points and integration scales from scale-space maxima (points that are local extrema with respect to both scale and space) of

$$\text{trace}\, \mathcal{H}_{norm} L = t\, (L_{xx} + L_{yy}), \quad \det \mathcal{H}_{norm} L = t^2\, (L_{xx} L_{yy} - L_{xy}^2), \tag{23}$$

where $\mathcal{H}_{norm} L$ denotes the normalized Hessian matrix of $L$. Then, given a point of interest and an integration scale $s$, the local scale is determined as described above.

[3] An intuitive motivation why this approach is suitable for blob-like image structures, is that the Laplacian operator can be expected to give a strong spatial response near the center of the blob, provided that the scale is tuned to the size of the blob. In addition, the scale, at which the maximum normalized Laplacian response over scales is assumed, can be expected to reflect the size of the blob.

The determinant of the Hessian has the additional advantage that it relative invariant under affine transformations. A disadvantage of $\det \mathcal{H}_{norm}$, however, is that this differential expression is more likely to give strong responses near corners.

# 4 Affine Gaussian scale-space

When dealing with linear transformations of the spatial domain, a natural generalization of the linear scale-space representation (based on the rotationally symmetric Gaussian kernel) is the *affine Gaussian scale-space representation* generated by convolution with non-uniform Gaussian kernels (Lindeberg 1994*c*). Given a symmetric positive semi-definite (covariance) matrix, $\Sigma_t \in \mathrm{SPSD}(2)$, the non-uniform Gaussian kernel in the two-dimensional case can be defined by

$$g(x;\ \Sigma_t) = \frac{1}{2\pi\sqrt{\det \Sigma_t}} e^{-x^T \Sigma_t^{-1} x/2}, \tag{24}$$

where $x \in \mathbb{R}^2$. In the special case when the matrix $\Sigma_t$ is a scalar entity $t$ times the unit matrix $I$ ($\Sigma_t = tI$), this function corresponds to the ordinary (uniform) Gaussian kernel with scale value $t$.

Given any function $f\colon \mathbb{R}^2 \to \mathbb{R}$, the affine Gaussian scale-space representation of $f$ can then be defined as the three-parameter family of functions $L\colon \mathbb{R}^2 \times \mathrm{SPSD}(2) \to \mathbb{R}$

$$L(\cdot;\ \Sigma_t) = g(\cdot;\ \Sigma_t) * f(\cdot). \tag{25}$$

This representation, considered in more detail in (Lindeberg 1994*a*), obeys the same scale-space properties as the linear scale-space representation, except those specifically connected to rotational symmetry. Moreover, due to the linearity, the scale-space properties transfer to spatial derivatives of the scale-space representation as well as to linear combinations of these. In this respect, the affine Gaussian scale-space representation constitutes the presumably simplest extension of the linear scale-space model to situations where further information is available.

## 4.1 Transformation property under linear transformations

A basic reason for introducing the affine Gaussian scale-space is that it is *closed* under linear (and affine) transformations of the spatial coordinates. Let $f_L, f_R\colon \mathbb{R}^2 \to \mathbb{R}$ be two intensity patterns related by an invertible linear transformation $\eta = B\xi$, i.e.,

$$f_L(\xi) = f_R(B\xi), \tag{26}$$

and define the affine Gaussian scale-space representations by

$$L(\cdot;\ \Sigma_L) = g(\cdot;\ \Sigma_L) * f_L(\cdot), \tag{27}$$

$$R(\cdot;\ \Sigma_R) = g(\cdot;\ \Sigma_R) * f_R(\cdot), \tag{28}$$

where $\Sigma_L, \Sigma_R \in \mathrm{SPSD}(2)$. Then, $L$ and $R$ are related by

$$L(\xi;\ \Sigma_L) = R(\eta;\ \Sigma_R), \tag{29}$$

where

$$\Sigma_R = B \Sigma_L B^T. \tag{30}$$

Hence, for any matrix $\Sigma_L$ there exists a matrix $\Sigma_R$ such that the affine Gaussian scale-space representations of $f_L$ and $f_R$ are equal (see the commutative diagram in figure 1). This property does not hold for the traditional linear scale-space representation based on the rotationally symmetric Gaussian (unless the linear transformation can be decomposed into rotations and uniform rescalings of the spatial coordinates).

$$
\begin{array}{ccccc}
L(\xi;\ \Sigma_L) & \text{---} & \eta = B\xi & \rightarrow & L(\eta;\ B\Sigma_L B^T) \\
\uparrow & & & & \uparrow \\
{*g}(\cdot;\ \Sigma_L) & & & & {*g}(\cdot;\ B\Sigma_L B^T) \\
| & & & & | \\
f_L(\xi) & \text{---} & \eta = B\xi & \rightarrow & f_R(\eta)
\end{array}
$$

Figure 1: Commutative diagram of the affine Gaussian scale-space representation under linear transformations of the spatial coordinates in the original image.

*Proof:* To verify the transformation property, insert $d(B\xi) = \det B \, d\xi$ into the (25):

$$
\begin{aligned}
L(q;\ \Sigma_L) &= \int_{\xi \in \mathbb{R}^2} g(q - \xi;\ \Sigma_L)\, f_L(\xi)\, d\xi \\
&= (\det B)^{-1} \int_{\xi \in \mathbb{R}^2} g(B^{-1}(Bq - B\xi);\ \Sigma_L)\, f_R(B\xi)\, d(B\xi).
\end{aligned}
\tag{31}
$$

Under a linear transformation, the non-uniform Gaussian kernel transforms as

$$
\begin{aligned}
g(B^{-1}\zeta;\ \Sigma_L) &= \frac{1}{2\pi\sqrt{\det \Sigma_L}}\, e^{-(B^{-1}\zeta)^T \Sigma_L^{-1} (B^{-1}\zeta)/2} \\
&= \sqrt{\det B B^T}\, \frac{1}{2\pi\sqrt{B \det \Sigma_L B^T}}\, e^{-\zeta^T (B\Sigma_L B^T)^{-1} \zeta/2},
\end{aligned}
\tag{32}
$$

which gives

$$
g(B^{-1}\zeta;\ \Sigma_L) = \det B \, g(\zeta;\ B\Sigma_L B^T). \tag{33}
$$

By inserting (33) in (31) with $\zeta = Bq - B\xi$ and by letting $\eta = B\xi$ with $p = Bq$ it follows that

$$
\begin{aligned}
L(q;\ \Sigma_L) &= \int_{\xi \in \mathbb{R}^2} g(Bq - B\xi;\ B\Sigma_L B^T)\, f_R(B\xi)\, d(B\xi) \\
&= \int_{\eta \in \mathbb{R}^2} g(Bq - \eta;\ B\Sigma_L B^T)\, f_R(\eta)\, d\eta \\
&= R(Bq;\ B\Sigma_L B^T) = R(p;\ \Sigma_R),
\end{aligned}
\tag{34}
$$

which proves (29) and (30).

## 4.2 Interpretation in terms of eigenvectors and eigenvalues

The effect of smoothing by convolution with the non-uniform Gaussian kernel can be easily understood in terms of the eigenvectors $\bar{b}$ and $\bar{t}$ and the eigenvalues $t_b \geq t_t > 0$ of $\Sigma_t^{-1}$. Let $u = (u_b, u_t)$ denote coordinates along the two eigenvectors respectively. The two coordinate systems are related by $u = Ux$, where $U$ is a unitary matrix ($UU^T = U^T U = I$) such that $\Sigma_t = U^T \Lambda_t U$ and $\Sigma_t^{-1} = U^T \Lambda_t^{-1} U$ , where $\Lambda_t$ is a diagonal matrix with $t_b$ and $t_t$ along the diagonal. In the transformed coordinates, the quadratic form can be written

$$
x^T \Sigma_t^{-1} x = x^T U^T \, \Lambda_t^{-1} \, U x = u^T \Lambda_t^{-1} \, u = \frac{u_b^2}{t_b} + \frac{u_t^2}{t_t}, \tag{35}
$$

which means that the non-uniform Gaussian kernel in this coordinate system assumes the form

$$g(u_b, u_t\ \Lambda_t) = \frac{1}{2\pi\sqrt{t_b t_t}} e^{-(u_b^2/2t_b + u_t^2/2t_t)} = \frac{1}{\sqrt{2\pi t_b}} e^{-u_b^2/2t_b}\ \frac{1}{\sqrt{2\pi t_t}} e^{-u_t^2/2t_t}. \tag{36}$$

In other words, convolution with (24) corresponds to (separable) smoothing with a one-dimensional Gaussian kernel with scale value $t_b$ along the $\bar{b}$-direction, and a one-dimensional Gaussian kernel with scale value $t_t$ along the $\bar{t}$-direction (see figure 2 for a few examples of such kernels).

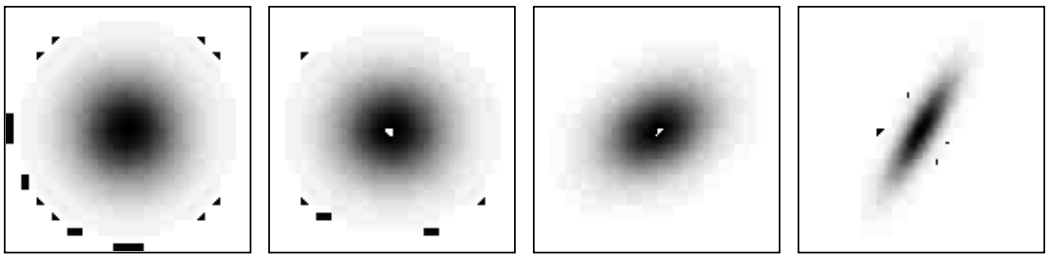

Figure 2: Examples of non-uniform Gaussian kernels. From left to right the eigendirection corresponding to the maximum eigenvalue is $(0°, 0°, 30°, 60°)$ and the minimum eigenvalue is decreased such that it corresponds to slant angles of $(0°, 22.5°, 45°, 77.5°)$.

### 4.3 Diffusion equation interpretation of affine Gaussian scale-space

Convolution with the non-uniform Gaussian kernel can also be interpreted in terms of the diffusion equation. Assume that $t_t$ and $t_b$ are related to a one-dimensional scale parameter $t$ by

$$t_b = t\alpha, \qquad t_t = t/\alpha, \tag{37}$$

for some constant[4] $\alpha > 1$. Then, in terms of the transformed coordinates, $(u_b, u_t)$, the non-uniform scale-space representation given by (25) satisfies the non-uniform diffusion equation

$$\partial_t L = \tfrac{1}{2}\,(\alpha\,\partial_{u_b u_b} + \frac{1}{\alpha}\,\partial_{u_t u_t})\,L \tag{38}$$

with initial condition $L(\cdot;\,0) = f$. With $\Lambda_t = t\Lambda_0$, where $\Lambda_0$ is a diagonal matrix with entries $\alpha$ and $1/\alpha$, and with $\nabla_u = (\partial_{u_b}, \partial_{u_t})^T$, the matrix form of this equation is

$$\partial_t L = \tfrac{1}{2}\,\nabla_u^T \Lambda_0 \nabla_u L. \tag{39}$$

In terms of the original coordinates, $(x, y)$, and with $\nabla = (\partial_x, \partial_y)^T$, the non-uniform diffusion equation assumes the form

$$\partial_t L = \tfrac{1}{2}\,\nabla^T \Sigma_0 \nabla L, \tag{40}$$

where $\Sigma_0 = U^T \Lambda_0 U$ is the positive definite matrix corresponding to $\Lambda_0$ rotated back to the $(x, y)$ coordinate system (i.e. $\Sigma_t = t\Sigma_0$). (Here, it has without loss of generality been assumed that the determinants of $\Lambda_0$ and $\Sigma_0$ are equal to one.)

[4] In the example in section 2 above, $\alpha$ corresponds to $1/\cos\sigma$.

### 4.4 Fourier transform of the non-uniform Gaussian kernel

From the Fourier transform of the one-dimensional Gaussian kernel

$$G(\omega;\ t) = \int_{x=-\infty}^{\infty} g(x;\ t)\, e^{-i\omega x} dx = e^{-\omega^2 t/2} \tag{41}$$

it directly follows that the Fourier transform of the non-uniform Gaussian kernel in the transformed $(u_b, u_t)$ coordinates is

$$G(w_b, w_t;\ \Lambda_t) = e^{-t_b w_b^2/2}\, e^{-t_b w_t^2/2} = e^{-(t_b w_b^2 + t_t w_t^2)/2} = e^{-w^T \Lambda_t w/2}.$$

With $w = U\omega$, and $\omega = (\omega_x, \omega_y)$ the Fourier transform in the original coordinate system can then be written

$$G(\omega;\ \Sigma_t) = e^{-\omega^T \Sigma_t \omega/2}. \tag{42}$$

From this expression it can be immediately realized that the semi-group property transfers to the non-uniform Gaussian kernel

$$G(\omega;\ \Sigma_1) * G(\omega;\ \Sigma_2) = G(\omega;\ \Sigma_1 + \Sigma_2). \tag{43}$$

## 5 Texture descriptors defined from affine Gaussian scale-space

We are now prepared to define a second moment matrix based on the affine Gaussian scale-space representation. Given an image $f\colon \mathbb{R}^2 \to \mathbb{R}$ with affine Gaussian scale-space representation $L\colon \mathbb{R}^2 \times \mathrm{SPSD}(2) \to \mathbb{R}$, the second moment matrix based on non-uniform smoothing $\mu_L\colon \mathbb{R}^2 \times \mathrm{SPSD}(2)^2 \to \mathrm{SPSD}(2)$ can be defined by

$$\mu_L(\cdot;\ \Sigma_t, \Sigma_s) = g(\cdot;\ \Sigma_s) * ((\nabla L)(\cdot;\ \Sigma_t)\, (\nabla L)(\cdot;\ \Sigma_t)^T) \tag{44}$$

where $\Sigma_s$ represents the covariance matrix corresponding to the integration scale, and $\Sigma_t$ the covariance matrix corresponding to the local scale.

### 5.1 Transformation property under linear transformations

Under a linear transformation of the image coordinates $\eta = B\xi$, this descriptor transforms as

$$\mu_L(q;\ \Sigma_t, \Sigma_s) = B^T \mu_R(Bq;\ B\Sigma_t B^T, B\Sigma_s B^T)\, B. \tag{45}$$

*Proof:* Differentiation of (29) gives

$$\nabla L(\xi;\ \Sigma_t) = B^T \nabla R(B\xi;\ B\Sigma_t B^T), \tag{46}$$

which implies that $\mu_L(q;\ \Sigma_t, \Sigma_s)$ assumes the form

$$\int_{\xi \in \mathbb{R}^2} g(q - \xi;\ \Sigma_s)\, B^T (\nabla R(B\xi;\ B\Sigma_t B^T))\, (\nabla R(B\xi;\ B\Sigma_t B^T))^T B\, d\xi.$$

After calculations similar to those in section 4.1, this expression can in turn be rewritten as

$$B^T \int_{\eta \in \mathbb{R}^2} g(Bq - \eta;\ B\Sigma_s B^T)\, (\nabla R(\eta;\ B\Sigma_t B^T))\, (\nabla R(\eta;\ B\Sigma_t B^T))^T\, d\eta\, B$$

which proves (45).

### 5.2 Shape adaptation: Invariance property of fixed points

With respect to the problem of shape adaptation, let us assume that at a certain image point we can compute the second moment matrix using shape adaptation such that the scale matrices are proportional to the inverse of the second moment matrix. Then, this fixed point will be preserved under affine transformations, and the backprojected window function (corresponding to (18)) transforms in the same way as the affine transformation.

To verify this property, consider a signal $f_L$, and assume that at a point $q_L \in \mathbb{R}^2$ the second moment matrix has been computed with shape adaptation[5] such that

$$\mu_L(q_L;\ \Sigma_{t,L}, \Sigma_{s,L}) = M_L, \tag{47}$$

where $\Sigma_{t,L}$ and $\Sigma_{s,L}$ are scalar multiples of $M_L^{-1}$,

$$\Sigma_{t,L} = t\, M_L^{-1}, \quad \Sigma_{s,L} = s\, M_L^{-1}, \quad (t, s \in \mathbb{R}_+). \tag{48}$$

Moreover, define a transformed intensity pattern $f_R$ by $f_L(\xi) = f_R(B\xi)$ for some linear transformation $\eta = B\xi$. Then, computation of the second moment matrix at a corresponding point in the transformed domain, $q_R = Bq_L$, using the same type of shape adaptation gives

$$\mu_R(q_R;\ \Sigma_{t,R}, \Sigma_{s,R}) = M_R, \tag{49}$$

where

$$\Sigma_{t,R} = t\, M_R^{-1}, \quad \Sigma_{s,R} = s\, M_R^{-1}. \tag{50}$$

*Proof:* Using (45) we obtain

$$\mu_L(q_L;\ \Sigma_{t,L}, \Sigma_{s,L}) = B^T \mu_R(q_R;\ B\Sigma_{t,L}B^T, B\Sigma_{s,L}B^T)\, B. \tag{51}$$

Hence,

$$M_R = \mu_R(q_R;\ \Sigma_{t,R}, \Sigma_{s,R}) = B^{-T} M_L B^{-1} \tag{52}$$

where

$$\Sigma_{t,R} = t\, B\Sigma_{t,L}B^T = t\,(BM_L^{-1}B^T) = t\,(B^{-T}M_L B^{-1})^{-1} = t\, M_R^{-1}. \tag{53}$$

A corresponding relation holds for $\Sigma_{s,R}$, which verifies the result.

In other words, if we can compute image descriptors that satisfy the fixed point conditions (47)–(48), these fixed point conditions will be preserved under arbitrary (invertible) affine transformations, and the image descriptors will be relative invariant under affine transformations.

In next section, it will be verified that this property transfers[6] to scale-space maxima of $\det \mu_L$. (Similar calculations show that it holds for $\det \mathcal{H}_{norm} L$ as well.)

[5] In certain situations it is useful to normalize the scale matrix for shape adaptation to have determinant one. Then, if $\Sigma_{t,L}$ and $\Sigma_{t,R}$ are rewritten as $\Sigma_{t,L} = t_L\, \Sigma_L$ and $\Sigma_{t,R} = t_R\, \Sigma_R$, where $\det \Sigma_L = \det \Sigma_R = 1$, $t_L$ and $t_R$ will be related by $t_R = t_L\, (\det B)^2$.

[6] This invariance property is non-trivial and does not hold, for example, for maxima over scales of $\operatorname{trace} \mu_{L,norm}$ or $\operatorname{trace} \mathcal{H}_{norm} L$.

### 5.3 Invariance property of normalized scale-space maxima

Assume that the shape and the orientation of the matrix used for shape adaptation is given and that only size variations are performed. Then, $\Sigma_t$ and $\Sigma_s$ can be written

$$\Sigma_t = t\Sigma_1, \quad \Sigma_s = s\Sigma_2, \tag{54}$$

for some *constant* matrices $\Sigma_1, \Sigma_2 \in \mathrm{SPSD}(2)$, and (45) assumes the form

$$\mu_L(q;\ t\Sigma_1, s\Sigma_2) = B^T \mu_R(Bq;\ tB\Sigma_1 B^T, sB\Sigma_2 B^T)\, B. \tag{55}$$

The corresponding expression for the (normalized) determinants becomes

$$\begin{aligned} \partial_t(t^2\ \det\Sigma_1\ \det\mu_L(q;\ t\Sigma_1, s\Sigma_2)) = 0 \quad \Leftrightarrow \\ \partial_t(t^2\ \det(B\Sigma_1 B^T)\ \det\mu_R(Bq;\ tB\Sigma_1 B^T, sB\Sigma_2 B^T)) = 0 \end{aligned} \tag{56}$$

showing that maxima over scales in the (normalized) determinant of the second moment matrix are preserved under linear transformations provided that the matrices $\Sigma_1$ and $\Sigma_2$ are matched accordingly. The same result applies to spatial maxima.

### 5.4 Interpretation of the invariance property in the case of weak isotropy

Selecting the shape adaptation matrix as the inverse of the second moment matrix has a simple geometric interpretation when estimating the affine deformation of a weakly isotropic pattern, *i.e.*, a pattern for which the second moment matrix is proportional to the unit matrix. Then,

$$M_R = cI \tag{57}$$

in (49), and at the fixed point it holds that the shape adaptation matrices $\Sigma_{t,R}$ and $\Sigma_{s,R}$ according to (50) are also proportional to the unit matrix.

In other words, for weakly isotropic patterns rotationally symmetric smoothing and rotationally symmetric window functions in the original domain (*e.g.*, in the case of shape from texture, the tangent plane to the surface) correspond to an affine invariant fixed point of the shape-adapted smoothing scheme.

## 6 Designing an iterative procedure

Given the fixed-point formulation of shape adaptation, it is natural to formulate an algorithm in terms of an iterative procedure. The invariance properties are then obtained provided that the procedure converges to the desired fixed point.

### 6.1 Variation of scale matrices

A general variation of the covariance matrices in the second moment matrix (44) based on the affine Gaussian scale-space representation obviously gives rise to a six-dimensional search space. With respect to the problem of finding the specific fixed point defined in section 5.2, this search space can be reduced by requiring the local and integration scale matrices to be coupled (according to (48)) such that

$$\Sigma_t = t\,\Sigma_0, \quad \Sigma_s = s\,\Sigma_0, \tag{58}$$

$$
\begin{array}{ccccc}
\mu_L(\xi;\ t(B^T B)^{-1}, s(B^T B)^{-1}) & & \eta = B\xi & & \mu_R(\eta;\ tI, sI) \\
= c' B^T B & \text{---} & \mu_L = B^T \mu_R B & \rightarrow & = c' I \\
\uparrow & & & & \uparrow \\
*g(\cdot;\ s(B^T B)^{-1}) & & & & *g(\cdot;\ sI) \\
\uparrow & & & & \uparrow \\
\nabla L_L \nabla L_L^T & & & & \nabla L_R \nabla L_R^T \\
| & & & & | \\
L_L(\xi;\ t(B^T B)^{-1}) & \text{---} & \eta = B\xi & \rightarrow & L_R(\eta;\ tI) \\
\uparrow & & & & \uparrow \\
*g(\cdot;\ t(B^T B)^{-1}) & & & & *g(\cdot;\ tI) \\
| & & & & | \\
f_L(\xi) & \text{---} & \eta = B\xi & \rightarrow & f_R(\eta)
\end{array}
$$

Figure 3: Interpretation of the shape-adapted smoothing operation in the case of weak isotropy. If the shape adaptation matrices are proportional to $(B^T B)^{-1}$, where $B$ describes the local linear backprojection from the image $f_L$ of a weakly isotropic pattern $f_R$ to the original domain, *i.e.*, $f_L(\xi) = f_R(B\xi)$, then, from the definition of the fixed-point condition, it follows that the estimate of the essential transformation properties (given by $\mu_L = cB^T B$) will (up to first order of approximation) be unaffected by the smoothing operation.

for some matrix $\Sigma_0$ (assumed to be normalized in some way; see next). This reduces the search space to four free parameters (two parameters determining the shape of $\Sigma_0$, and the two other parameters determining the size of each scale matrix).

Here, we shall consider the following methods for choosing these parameters in an iterative shape adaptation scheme (see section 10 for further discussion):

**Shape of the smoothing matrices.** Concerning the choice of the first two parameters, we shall here adopt the straightforward approach of letting the shape of $\Sigma_0$ be *proportional to the second moment matrix* at the given point in each iteration.

**Size of the integration scale matrix.** To preserve invariance under size variations it is natural to require the size of the integration scale matrix to be selected such that the *normalized differential entity* considered for scale selection ($\det \mu_{L,norm}$ or $\det \mathcal{H}_{norm} L$) one may require) still *assumes a maximum over scales* (compare with section 5.3). If the size variations are small it is also reasonable to keep $s$ constant (for example, for binocular vision with approximately symmetric vergence).

**Size of the local scale matrix.** Some further precautions[7] must be taken when choosing the size of local scale matrix. It is clear that selecting it such that the normalized anisotropy $\tilde{Q}$ is still a maximum over scales is not an appropriate solution in the non-isotropic case.[8] To formulate an alternative criterion, one may at first find

[7] Note that invariance arguments cannot be applied here, since a main purpose of the local smoothing is to suppress noise, which is an image effect decoupled from the surface geometry.

[8] The motivation for selecting this scale was that in the rotationally symmetric case, local smoothing could be expected to increase the anisotropy when it serves the purpose of suppressing noise,

it natural to normalize $\Sigma_0$ to $\det \Sigma_0 = 1$ and then assume that $t$ should be constant during the iterations. Such an approach would, however, also lead to systematic overestimates of the anisotropy and, hence, the slant angle. This can be easily understood by observing that the amount of smoothing in the direction of the smallest eigenvalue of the local scale matrix will then be smaller in the non-isotropic case than in the isotropic case. Hence, the contribution from the fine scale variations in the signal (here, the noise) can be expected to be larger, which means that the anisotropy will increase.

Here, given an initial estimate based on a maximum over scales of $\tilde{Q}$ (computed with rotationally symmetric smoothing), we propose to keep the *smallest eigenvalue* $\lambda_{min}(\Sigma_t)$ of $\Sigma_t$ constant during the iterations. By this approach, the smallest amount of smoothing will still be preserved, and the same (minimum) degree of noise suppression is guaranteed.

### 6.2 Composed method for shape adaptation

To summarize, a straightforward iterative method for shape adaptation can be expressed as follows ($M_I^{(k)}$ denotes the matrix for shape adaptation in the $k$th iteration step, $\mathcal{D}_{norm}L$ represents the normalized differential entity used for selecting the integration scale, and $s_{\mathcal{D}L}^{(k)}$ the selected integration scale in each step):

1. $M^{(0)} = I$.
2. $M_I^{(k)} = M^{(k)} / \lambda_{min}(M^{(k)})$.
3. $s_{\mathcal{D}L}^{(k)} = \begin{cases} \gamma^2 \cdot (s\colon \max_{s>0} (\mathcal{D}_{norm}L)(q;\ sM_I^{(k)})), \\ \text{alternatively, let } s_{\mathcal{D}L}^{(k)} = s_{\mathcal{D}L}^{(0)} \text{ for all } k \geq 1. \end{cases}$
4. $M^{(k+1)} = \mu_L^{-1}(q;\ t\, M_I^{(k)},\ s_{\mathcal{D}L}^{(k)} M_I^{(k)})$. Go to step 2 if not converged.

When formalizing a termination criterion, it is natural to require that successive second moment matrices should be sufficiently "similar". When formulating such a similarity measure for a multi-dimensional entity some precautions need to be taken. A natural approach in this case is to represent each matrix $\mu$ by an ellipse $x^T \mu\, x = 1$, interpreting each ellipse as the orthographic projection of a planar circle, and measuring the angle between the three-dimensional orientations of the circles. Whereas for weakly isotropic surface patterns this measure is equal to the angle between the surface normals, this measure is a natural choice also in other situations.

In summary, a useful convergence criterion is to require this angular difference between successive second moment matrices to be less than some given threshold $\Delta\varphi < \varepsilon$, and to impose an upper bound on the number of iterations.

## 7 Estimating surface orientation: Review

This section reviews a previously developed methodology for estimating surface orientation from either monocular texture distortion or the gradient of binocular disparity. These methods are based on measurements of affine transformations using second moment matrices, *i.e.* the same entities as we here use for shape adaptation.

whereas too much local smoothing can be expected to decrease the anisotropy when it leads to shape distortions. Hence, selecting the maximum gives a natural trade-off between these two effects.

## 7.1 Shape from texture

The systematic distortion of texture patterns under perspective transformations can under certain conditions provide strong cues to surface orientation and surface shape. Whereas the general problem of determining three-dimensional surface structure from monocular data is obviously underdetermined (because of dimensionality considerations), it may become feasible if additional *a priori* information is available. This is the subject of *shape-from-texture*: to combine measurements of image structures with different assumptions about the surface texture and the camera geometry to derive three-dimensional shape cues.

This idea of using perspective distortion of texture as a cue to surface shape goes back to Gibson (1950), who coined the term "texture gradient" for the systematic variation of image textures induced by perspective effects. Since then, computational studies of the shape-from-texture problem have been done by several researchers, for example, (Bajcsy & Lieberman 1976, Stevens 1981, Witkin 1981, Kanatani 1984, Pentland 1986, Aloimonos 1988, Blostein & Ahuja 1989, Kanatani & Chou 1989, Blake & Marinos 1990*b*, Blake & Marinos 1990*a*, Brown & Shvaytser 1990, Gårding 1992, Gårding 1993, Lindeberg & Gårding 1993, Malik & Rosenholtz 1993). Here, we shall consider the shape-from-texture method by (Lindeberg & Gårding 1993), which will be briefly reviewed.

### 7.1.1 Shape-from-texture using local affine distortion

With respect to the shape-from-texture problem, let the linear transformation $B$ represent the derivative of the mapping $A\colon \Pi \to S$ from the planar image $\Pi$ onto the surface $S$. In practice, it is convenient to decompose $A = F \circ G$ into the so-called gaze-transformation $G\colon \Pi, q \to \Sigma, p$ from the planar image to the viewsphere $\Sigma$, and the perspective backprojection $F\colon \Sigma, p \to S, F(p)$ from the viewsphere to the surface. The chain rule of differentiation gives $B = A_{*q} = F_{*p}\, G_{*q}$, where $F_{*p}$ and $G_{*q}$ denote the derivatives[9] of $F$ and $G$ respectively.

To relate local surface orientation to these entities, introduce a local coordinate system on the viewsphere such that for any point $p \in \Sigma$ the triple $(\bar{p}, \bar{t}, \bar{b})$ is a local orthonormal coordinate system with $\bar{p}$ as view direction. Define the *tilt direction* $\bar{t}$ as the direction of the gradient of the distance from the focal point, and let $\bar{b} = \bar{p} \times \bar{t}$. Moreover, in $T_{F(p)}(S)$, let $\bar{T}$ and $\bar{B}$ be the normalized images of $\bar{t}$ and $\bar{b}$ respectively. From the differential geometric analysis in (Gårding 1992) it follows that in the bases $(\bar{t}, \bar{b})$ and $(\bar{T}, \bar{B})$ the expression for $F_{*p}$ is

$$F_{*p} = \begin{pmatrix} r/\cos\sigma & 0 \\ 0 & r \end{pmatrix} = \begin{pmatrix} 1/m & 0 \\ 0 & 1/M \end{pmatrix}, \tag{59}$$

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), $\sigma$ is the slant angle, and $m < M$ are the inverse eigenvalues of $F_{*p}$.

This framework describes how local surface orientation can be related to local affine distortion in the monocular case. With $B = A_{*q}$ (17) gives

$$\mu_L(q) = G_{*q}^T\, F_{*p}^T\, \mu_S(F(G(q)))\, F_{*p}\, G_{*q}, \tag{60}$$

[9] Using traditional differential geometric notation, $F_{*p}\colon T_p(\Sigma) \to T_{F(p)}(S)$ is the derivative of $F$ at $p = G(q) \in \Sigma$, $G_{*q}\colon T_q(\Pi) \to T_p(\Sigma)$ is the derivative of $G$ at $q \in \Pi = \mathbb{R}^2$, and $T_p(\Sigma)$ is the tangent plane of $\Sigma$ at $p$, etc.

where $\mu_S(F(G(q)))$ denotes the second moment matrix defined in the tangent plane to the surface with respect to the window function

$$w'(\eta - F(G(q))) = (\det A_{*q})^{-1} w(\xi - q) \tag{61}$$

where $\eta = A_{*q}\xi$. A 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 $\mu_S(F(G(q)))$ to infer the structure of $A_{*q}$. This permits computation of $F_{*p}$ after compensation with respect to $G_{*q}$ (which is known if the camera geometry is).

The assumption that leads to the simplest estimation procedure is *weak isotropy*, *i.e.*, that $\mu_S(F(p))$ is proportional to the unit matrix ($\tilde{Q} = 0$). Under this condition and if $F_{*p}$ is non-degenerate, it holds that the tilt direction is parallel to the eigenvalue of $\mu_L$ corresponding to the maximum eigenvalue and

$$\cos\sigma = \frac{m}{M} = \sqrt{\frac{\lambda_2}{\lambda_1}} = \sqrt{\frac{1-\tilde{Q}}{1+\tilde{Q}}}, \tag{62}$$

which gives a direct estimate of local surface orientation.

This scheme can be applied also under other assumptions. For example, if the local "size" of the surface texture elements does not vary systematically, then $A = 1/\sqrt{\det \mu}$ is an area measure, and the normalized *area gradient* provides information about both local surface orientation and curvature (Gårding 1992, Lindeberg & Gårding 1993, Gårding & Lindeberg 1996). An alternative assumption is that the surface texture can be well modelled as the realization of a stationary process and that the spectrogram can be assumed to be approximately constant when measured in different windows over the surface (Malik & Rosenholtz 1993).

### 7.2 Shape-from-disparity-gradients

A similar approach can be applied to binocular data. If we have two cameras verging onto the same surface structure, then we can make use of the fact that two different observations are made of the same surface structure. If the mapping between the left and the right image is approximated by an affine transformation, and if this component of the perspective transformation can be measured, then the surface orientation can be computed provided that the camera orientation is known. The main difference compared to the monocular case is that no specific assumptions about the surface texture will be necessary.

Several computational models have been described in the literature that are based more or less directly on this type of cue, *e.g.*, (Blakemore 1970, Koenderink & van Doorn 1976, Tyler & Sutter 1979, Rogers & Cagenello 1989, Wildes 1981, Jones & Malik 1992). Here, we shall consider a shape-from-disparity-gradient method which is a straightforward extension of the previous shape-from-texture method (Gårding & Lindeberg 1994, Gårding & Lindeberg 1996). It assumes epipolar geometry and verging cameras,[10] and computes a surface orientation estimate at the fixation point.

[10] In the work by (Jones & Malik 1992) symmetric vergence was assumed. This assumption is, however, not necessary for the analysis. A generalization to the case of asymmetric vergence is given in (Gårding & Lindeberg 1994).

Moreover, there is no need to assume that the cameras verge *physically* in these models. All that is required is that the camera geometry is known (calibrated stereo) such that compensations (virtual camera movements or virtual image warpings) can be performed.

### 7.2.1 Shape-from-disparity-gradients using local affine distortion

Following (Jones & Malik 1992) consider an affine transformation of the mapping from the left to the right camera

$$\begin{pmatrix} x_R \\ y_R \end{pmatrix} = \begin{pmatrix} 1 + H_x & H_y \\ V_x & 1 + V_y \end{pmatrix} \begin{pmatrix} x_L \\ y_L \end{pmatrix} + \begin{pmatrix} H \\ V \end{pmatrix}. \tag{63}$$

Since the cameras are assumed to verge onto the same physical structure, the horizontal disparity $H$ and the vertical disparity $V$ are zero. Moreover, due to the assumption about epipolar geometry, the components of the vertical disparity gradient, $V_x$ and $V_y$, are zero. According to the analysis in (Jones & Malik 1992), the components of the horizontal disparity gradient can then be related to local surface geometry by

$$\begin{aligned} \tan \phi_y &= \frac{H_x}{H_x + 2} \cot \delta, \\ \tan \phi_x &= \frac{-H_y}{\sqrt{H_x^2 + 4(H_x + 1) \sin^2 \delta}}, \end{aligned} \tag{64}$$

where $2\delta$ is the vergence angle, and the orientation of the surface relative to the cyclopean image plane is represented by a rotation $\phi_x$ around the $x$-axis followed by a rotation $\phi_y$ around the $y$-axis. These angles are related to slant $\sigma$ and tilt $\theta$ by

$$\cos \sigma = \cos \phi_x \cos \phi_y, \quad \tan \theta = -\tan \phi_x / \sin \phi_y. \tag{65}$$

Assume now that second moment matrices $\mu_L$ and $\mu_R$ can be measured in the left and the right images, and let $B$ represent the linear transformation (63). Then, according to (17), the second moment matrices are related by

$$\mu_L(q) = B^T \mu_R(p) B. \tag{66}$$

This relation imposes three restrictions onto the two unknowns, $H_x$ and $H_y$. One way to relax the overdeterminacy (Gårding & Lindeberg 1996) is to consider only the directional information in $\mu_L$ and $\mu_R$ and to ignore the size information. After some algebraic manipulations it can then be shown that the horizontal disparity gradient can be computed from the components of the second moment matrices using

$$\begin{aligned} H_x &= \frac{1 + \tilde{C}_L}{1 + \tilde{C}_R} \sqrt{\frac{1 - \tilde{Q}_R^2}{1 - \tilde{Q}_L^2}} - 1, \\ H_y &= \frac{\tilde{S}_L \sqrt{1 - \tilde{Q}_R^2} - \tilde{S}_R \sqrt{1 - \tilde{Q}_L^2}}{(1 + \tilde{C}_R) \sqrt{1 - \tilde{Q}_L^2}}, \end{aligned} \tag{67}$$

where for each second moment matrix the entities $\tilde{C}$ and $\tilde{S}$ are defined as

$$P = E_q(L_x^2 + L_y^2), \quad \tilde{C} = E_q(L_x^2 - L_y^2)/P, \quad \tilde{S} = E_q(2L_x L_y)/P. \tag{68}$$

The main difference between this method and the work by (Jones & Malik 1992) is that here *direct estimates* are obtained and no explicit search is necessary.

# 8 Experiments

The proposed scheme for shape adaptation has been applied to the problem of estimating surface orientation using the methods reviewed in previous section. Experiments have been performed on real and synthetic reference images with known orientation. To test the stability of the method, Gaussian noise of different standard deviation has been added to the images. Some test images are shown in figure 4.

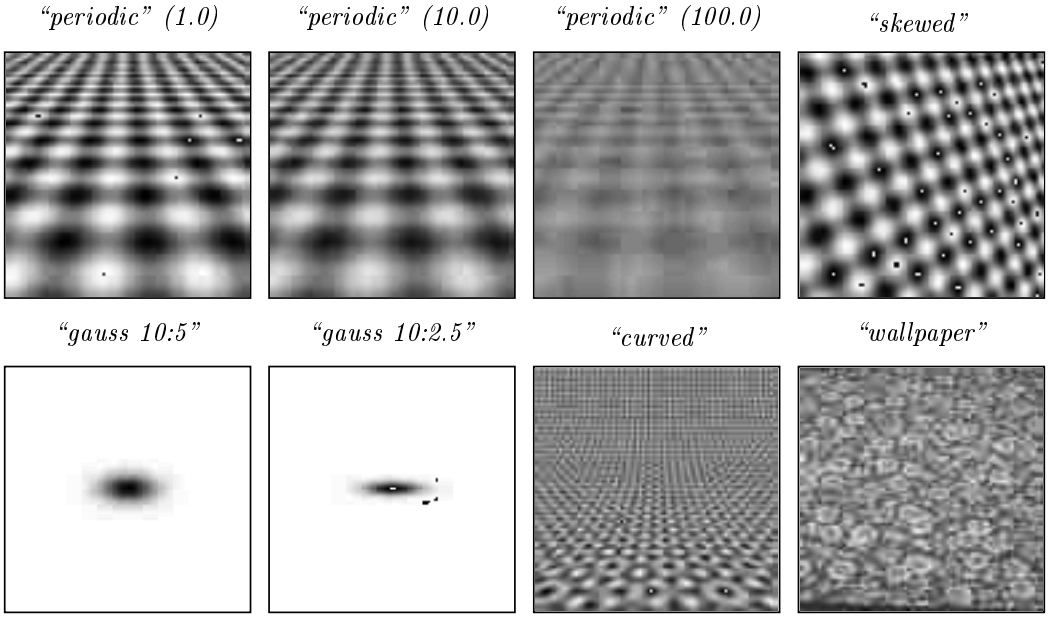


Figure 4: Grey-level images used in the experiments on shape-from-texture. The first image, called 'periodic', is shown with added Gaussian noise of standard deviation 1.0, 10.0, and 100.0, respectively. while for the other test data only the original image is shown.

## 8.1 Shape-from-texture

Table 1 shows the result of applying the scheme to the image labelled "periodic" and using different amounts of added noise (the standard deviation is shown in parenthesis in the header of each table). The columns in each table show from left to right, the iteration index, surface orientation estimates in terms of slant and tilt angles computed under the assumption of weak isotropy, and the angle between the estimated direction and the true direction of the surface normal. To compare the effect of shape adaptation with the effect of increasing the integration scale in the isotropic scale-space representation, experimental results are also given using different values of the *relative integration scale*, $\gamma$.[11] Notice that for high noise levels the shape adaptation leads to improvements in estimated surface orientation that cannot be achieved by

[11] This parameter (defined in section 3.2) determines how large the integration scale is relative to the scale at which the maximum over scales in the scale-space signature is assumed. Basically, a larger value of $\gamma$ can be expected to increase the accuracy up to the point where the modelling error increases due to violation of the local linear approximation.

just increasing the integration scale. Moreover, observe that substantial improvements can be obtained after just one or two iterations.

| periodic (1.0) ($\gamma = 1.0$) | | |
|---|---|---|
| 0 | (56.94, 80.94) | 8.29 |
| 1 | (60.03, 86.52) | 3.00 |
| 2 | (60.10, 87.55) | 2.12 |
| 3 | (60.09, 87.61) | 2.07 |
| 4 | (60.09, 87.61) | 2.07 |

| periodic (10.0) ($\gamma = 1.0$) | | |
|---|---|---|
| 0 | (56.31, 81.13) | 8.37 |
| 1 | (60.18, 86.32) | 3.18 |
| 2 | (60.42, 87.55) | 2.16 |
| 3 | (60.42, 87.67) | 2.06 |
| 4 | (60.41, 87.67) | 2.05 |

| periodic (100.0) ($\gamma = 1.0$) | | |
|---|---|---|
| 0 | (47.11, 74.45) | 17.90 |
| 1 | (54.18, 81.32) | 9.31 |
| 2 | (56.39, 84.21) | 6.09 |
| 3 | (57.03, 85.33) | 4.96 |
| 4 | (57.21, 85.69) | 4.61 |

| periodic (1.0) ($\gamma = 1.4$) | | |
|---|---|---|
| 0 | (58.88, 89.22) | 1.29 |
| 1 | (60.08, 90.12) | 0.13 |
| 2 | (60.09, 90.11) | 0.13 |
| 3 | (60.08, 90.11) | 0.13 |
| 4 | (60.08, 90.11) | 0.13 |

| periodic (10.0) ($\gamma = 1.4$) | | |
|---|---|---|
| 0 | (57.88, 88.95) | 2.29 |
| 1 | (60.51, 90.09) | 0.52 |
| 2 | (60.75, 90.11) | 0.76 |
| 3 | (60.77, 90.10) | 0.77 |
| 4 | (60.77, 90.10) | 0.77 |

| periodic (100.0) ($\gamma = 1.4$) | | |
|---|---|---|
| 0 | (53.86, 89.61) | 6.14 |
| 1 | (59.90, 90.76) | 0.67 |
| 2 | (61.45, 90.90) | 1.65 |
| 3 | (61.94, 90.89) | 2.09 |
| 4 | (62.10, 90.88) | 2.24 |

| periodic (1.0) ($\gamma = 2.0$) | | |
|---|---|---|
| 0 | (60.29, 90.18) | 0.33 |
| 1 | (60.59, 89.95) | 0.59 |
| 2 | (60.59, 89.95) | 0.60 |
| 3 | (60.59, 89.95) | 0.60 |
| 4 | (60.59, 89.95) | 0.60 |

| periodic (10.0) ($\gamma = 2.0$) | | |
|---|---|---|
| 0 | (58.80, 90.04) | 1.19 |
| 1 | (60.49, 89.90) | 0.49 |
| 2 | (60.66, 89.90) | 0.66 |
| 3 | (60.68, 89.90) | 0.68 |
| 4 | (60.68, 89.90) | 0.68 |

| periodic (100.0) ($\gamma = 2.0$) | | |
|---|---|---|
| 0 | (49.95, 91.24) | 10.10 |
| 1 | (56.72, 90.74) | 3.34 |
| 2 | (58.38, 90.40) | 1.65 |
| 3 | (58.87, 90.32) | 1.16 |
| 4 | (59.02, 90.30) | 1.00 |

Table 1: Experiments with shape adapted smoothing applied on the image labelled "periodic" (using different values of the relative integration scale $\gamma$). The columns show from left to right, the iteration index, the slant and tilt values (reference values (60.0, 90.0)), and the angle between the estimated and the true surface normal. (The noise level gives the standard deviation of the noise, to be related to the grey-level range [0, 255]. All angles in degrees.)

Corresponding experimental results for the Gaussian blobs and the three other images are given in table 2 and table 3 respectively To save space, the results are shown for a smaller number of combinations of noise levels and values of $\gamma$, and only the first two iterations are displayed.

## 8.2 Shape-from-disparity-gradients

The need for shape adaptation can be further motivated when dealing with binocular data, since multiple measurements are made of the same surface structure, and the difference between these measurements is used as the basis for inferring cues to the three-dimensional surface structure. Hence, when non-negligible amounts of smoothing are required (typically, in the presence of noise or other interfering fine-scale structures), substantial improvements can be expected in the accuracy of the resulting surface orientation estimates. Moreover, it is desirable if the support regions of the image operators in the different images correspond to the same region when backprojected onto the surface.

Table 4 shows the results of including shape adaptation as an essential step in the shape-from-disparity-gradient method reviewed in section 7.2 and applying it to the stereo pair in figure 5. (The shape adaptation has been performed on both images independently before the surface orientation estimates have been computed from

| gauss 10:5 (1.0) | | |
|---|---|---|
| 0 | (66.71, 90.04) | 6.71 |
| 1 | (58.83, 90.01) | 1.16 |
| 2 | (60.10, 90.02) | 0.10 |

| gauss 10:5 (10.0) | | |
|---|---|---|
| 0 | (65.55, 89.52) | 5.56 |
| 1 | (58.66, 89.81) | 1.34 |
| 2 | (59.54, 89.73) | 0.50 |

| gauss 10:5 (100.0) | | |
|---|---|---|
| 0 | (66.71, 92.13) | 6.98 |
| 1 | (62.13, 90.64) | 2.20 |
| 2 | (61.72, 90.68) | 1.82 |

| gauss 10:2.5 (3.1) | | |
|---|---|---|
| 0 | (80.68, 89.93) | 5.16 |
| 1 | (75.99, 90.00) | 0.47 |
| 2 | (75.26, 90.00) | 0.25 |

| gauss 10:2.5 (10.0) | | |
|---|---|---|
| 0 | (80.24, 89.90) | 4.72 |
| 1 | (75.89, 89.95) | 0.37 |
| 2 | (75.25, 89.95) | 0.27 |

| gauss 10:2.5 (31.6) | | |
|---|---|---|
| 0 | (79.36, 90.28) | 3.85 |
| 1 | (76.91, 90.17) | 1.39 |
| 2 | (75.93, 90.03) | 0.41 |

Table 2: Shape adaptation applied to two Gaussian blobs using different amounts of added white Gaussian noise. The reference orientation in the top row is (60.0, 90.0) and in the bottom row (75.5, 90.0). (In these experiments the relative integration scale is $\gamma = 1.0$.)

| skewed (10.0) ($\gamma = 1.4$) | | |
|---|---|---|
| 0 | (28.81, 20.91) | 1.26 |
| 1 | (29.71, 20.95) | 0.55 |
| 2 | (29.75, 20.95) | 0.53 |

| curved (10.0) ($\gamma = 1.4$) | | |
|---|---|---|
| 0 | (59.57, 92.07) | 4.89 |
| 1 | (55.99, 90.61) | 1.11 |
| 2 | (56.22, 90.73) | 1.36 |

| wallpaper (10.0) ($\gamma = 16$) | | |
|---|---|---|
| 0 | (46.09, 85.56) | 4.71 |
| 1 | (51.86, 85.44) | 1.06 |
| 2 | (53.74, 85.34) | 2.94 |

Table 3: Shape adaptation applied to the images labelled "skewed", "curved", and "wallpaper". The reference orientations are (30.0, 20.0), (55.0, 90.0), and (50.8, 85.3) respectively.

(??)–(67) and (??)–(64).) Observe how the error (measured as the angle between the estimated and the true directions of the surface normal) decreases with the iterations.

*"skewed-R (10.0)"* *"skewed-L (10.0)"*

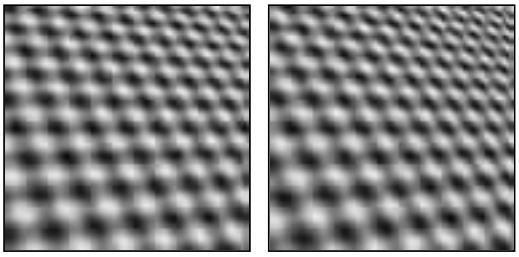

Figure 5: A stereo pair used in the experiments on shape-from-disparity-gradients (arranged for cross-eyed fusion). (The standard deviation for the added Gaussian noise is 10.0.)

An obvious alternative approach is to let the shape adaptation procedure be driven by the inter-image deformations. This corresponds to adjusting the shapes of (one or both of) the Gaussian kernels such that they (to first-order of approximation) backproject to the same surface patch in the world (see figure 6).[12]

[12]The affine transformation $B$ from $\mu_L$ and $\mu_R$ can the be determined either from the general expression (14) or the explicit expressions in section 6.2.2 in (Gårding & Lindeberg 1996), which make use of the fact that the relative orientation states between the cameras are known.

| skewed (10.0) ($\gamma = 1.4$) | | | skewed (10.0) ($\gamma = 2.0$) | | | skewed (10.0) ($\gamma = 2.8$) | | |
|---|---|---|---|---|---|---|---|---|
| 0 | (50.15, 65.23) | 5.99 | 0 | (47.69, 70.04) | 10.33 | 0 | (49.53, 68.77) | 8.47 |
| 1 | (54.19, 62.71) | 2.12 | 1 | (50.24, 64.62) | 5.61 | 1 | (52.05, 63.53) | 3.72 |
| 2 | (54.70, 61.31) | 0.94 | 2 | (51.72, 62.47) | 3.42 | 2 | (53.32, 61.87) | 1.89 |
| 3 | (54.96, 61.11) | 0.85 | 3 | (52.52, 61.43) | 2.32 | 3 | (53.88, 61.12) | 1.07 |
| 4 | (55.02, 61.01) | 0.81 | 4 | (52.94, 60.89) | 1.77 | 4 | (54.18, 60.80) | 0.67 |
| 5 | (55.04, 60.99) | 0.80 | 5 | (53.19 ,60.58) | 1.45 | 5 | (54.33, 60.61) | 0.46 |

Table 4: Shape adaptation applied to the stereo images in figure 5. The vergence angle is $2\delta = 10.00$ and the reference orientation with respect to a cyclopean coordinate system is (54.60, 60.16) (corresponding to $\phi_x = -45.00$ and $\phi_y = 35.00$).

# 9 Summary

We have described how shape distortions may arise when computing three-dimensional shape cues from image data using operators (derivatives) defined from linear scale-space representation. A methodology for reducing these problems has been presented, based on a straightforward extension of the linear scale-space concept into an affine Gaussian scale-space representation generated by convolution with non-uniform Gaussian kernels. The suggested approach is to adapt the shape of the smoothing kernel to the local image structure by measuring an image descriptor called the second moment matrix. This descriptor represents statistics of gradient directions in a neighbourhood of the point at which the descriptor is defined, and is a useful entity for measuring local affine distortions in shape-from-X algorithms.

If shape adaptation can be performed such that the second moment matrix computed at a certain point is proportional to the matrix used for shape adaptation at that point, then this fixed point will be preserved under affine transformations of the brightness pattern. In other words, the image descriptors will be relative invariant under the locally linearized perspective mapping, and (to first-order of approximation) the smoothing operation corresponds to using the same smoothing kernel in the tangent plane to the surface—independent of the view direction and independent of the shape-from-X method that uses the data.

In the shape-from-texture case, the shape adaptation means that the errors due to the smoothing operation will be confined to the higher order terms in the linear approximation, which in turn means that the accuracy in the computed surface orientation estimates can be expected to increase.

Similar properties hold for shape estimation from disparity gradients provided that the iterative scheme converges to corresponding fixed points in the left and right images. In this case, the geometric interpretation of the shape adaptation scheme is that the image operators in the left and right images correspond to similar operations when backprojected onto the image domain. Hence, the actual effects of the Gaussian derivative operators are the same and the window functions in the left and right images correspond to the same region on the surface.

Whereas the technical treatment in this article has been concerned with shape adaptation for the specific schemes for shape-from-texture and shape-from-disparity-gradients, the underlying ideas are of much wider generality. The notion of shape adaptation is essential to any method for estimating three-dimensional shape cues from measurements of affine distortions of two-dimensional brightness structures over regions of non-infinitesimal size. Without this operation, it is impossible to achieve

exact equality in the affine transformation between images of the same surface patch taken from different view points.[13] Therefore there will be systematic errors in the estimates, since the measurements performed in the different images do not correspond to the same physical region when backprojected to the surface. Besides shape from texture and stereo cues, this problem arises also in motion estimation and structure from motion. How large these errors become in practice is determined by the relation between the structures covered by registrations in both domains and the structures covered just by one (see figure 6 for an illustration).

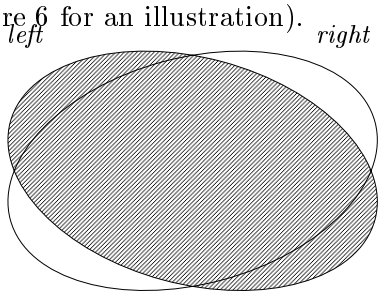


Figure 6: If a method for shape estimation from disparity gradients is based on rotationally symmetric operators in the image domain, then the backprojections of the operators from the left and right images to the physical domain will in general, correspond to different regions, in which the information is weighted differently. Hence, if the difference between these two measurements is used as a basis for estimating surface shape (for example, using the type of methodology described in section 7.2), the shape estimates will, in general, be biased unless that shapes of the operators are adapted such that the backprojections correspond. How large the bias becomes is in practice is, of course, strongly dependent on the image data.

When estimating local surface orientation, first-order shape adaptation is sufficient to obtain estimates that are unaffected by the smoothing operation up to the same order of approximation as the geometric approximation. To allow for unbiased curvature estimation, higher-order shape adaptation will then, in general, be needed.

# 10 Discussion

For the first-order shape adaptation approach based on second moment matrices, a straightforward algorithm has been presented for reaching the fixed point. Whereas it in the experiments has lead to substantial improvements after just a few iterations, no claims are made that it constitutes any "optimal solution". (Compare with the vast number of iterative methods for solving non-linear equations in numerical analysis and optimization theory.) We are currently studying convergence properties in more detail as well as the ability to reach appropriate fixed points. As a brief indication of the convergence properties, it can be mentioned that for a periodic pattern

$$f(x, y) = \cos \omega_1 x + \cos \omega_2 y \qquad (\omega_1 < \omega_2), \tag{69}$$

[13] For the specific shape-from-X method presented here, based on the second moment matrix, the violation of the linear transformation property manifests itself as follows: If measurements of second moment matrices $\mu_L$ and $\mu_R$ in both domains are performed based on rotationally symmetric Gaussian kernels, then the transformation property in (17) is not valid (although in many cases it may constitute a reasonable approximation). The transformation property holds (exactly) only if the shapes of the Gaussian kernels are modified, as expressed in (45).

the desired fixed point is convergent if the minimum amount of smoothing $t_0$ satisfies

$$\omega_1^2 t_0 < (\omega_1/\omega_2)^2 \tag{70}$$

(see appendix A). For example, for most images, there are (at least) two additional fixed points; one corresponding to the shape adaptation matrix being singular[14] ($\tilde{Q} = 1$, i.e., $\sigma = \pi/2$ in the case of weak isotropy) and one corresponding to an infinite amount of isotropic smoothing ($\tilde{Q} = 0$, i.e., $\sigma = 0$ in the case of weak isotropy).

During recent years, a large number of approaches to non-uniform smoothing have been proposed in the literature based on different non-linear versions of the diffusion equation (see section 1.2). The suggested scheme has a number of interesting relationships to these. If applied at edge points, it leads to larger amount of smoothing in the direction along the edge than in the perpendicular direction (compare with the closely related work by (Nitzberg & Shiota 1992)). Moreover, the non-linear diffusion equation (3) closely relates to a singular limit case of this method (when $\tilde{Q} = 1$).

Since the proposed shape adaptation method is based on rotationally symmetric smoothing as an uncommitted first processing step, which is then used for invoking more refined processing, the suggested scheme provides an interesting connection between the linear and non-linear scale-space approaches—potentially also as a step towards a common framework for unifying processing modules based on sparse edge data and dense filter outputs.

# A Convergence analysis

This section analyses the convergence properties of the iterative shape adaptation algorithm in a specific case. (Another example is analysed in section 2.)

## A.1 Convergence analysis for a periodic signal

**Model.** Consider a two-dimensional signal defined as the sum of two cosine waves along the two coordinate directions respectively,

$$f(x, y) = \cos \omega_1 x + \cos \omega_2 y \quad (\omega_1 \leq \omega_2). \tag{71}$$

This function can be interpreted as a simple locally linearized model of the projection of a periodic pattern. The ratio

$$\epsilon = \frac{\omega_1}{\omega_2} = \cos \sigma \tag{72}$$

describes the foreshortening, where $\sigma$ denotes the slant angle.

[14] If the shape adaptation procedure reaches a singular fixed point with $\tilde{Q} = 1$, this is a strong indicator that there is only one dominant orientation in the support region of the window function $w$, and that the surface orientation estimate, hence, cannot be trusted.

In a related work, (Almansa & Lindeberg 1996) apply a closely related shape adaptation procedure to a dense set of points. To avoid that the filters become too elongated (an that the shape adaptation process approaches a singular fixed point), an upper bound is introduced on the eccentricity of the filters by letting the shape adaptation matrix be given by $\Sigma = (\mu + \varepsilon I)^{-1}$, where $\varepsilon = \frac{Q}{\kappa - 1}$ and $\kappa$ serves as an upper bound on the condition number of $\Sigma$.

**Scale-space representation.** The scale-space representation of $f$ is given by

$$L(x, y;\ t) = e^{-\omega_1^2 t/2} \cos \omega_1 x + e^{-\omega_2^2 t/2} \cos \omega_2 y. \tag{73}$$

Differentiation gives the scale-space derivatives

$$\begin{aligned} L_x(x, y;\ t) &= -\omega_1 e^{-\omega_1^2 t/2} \sin(\omega_1 x + \phi_1), \\ L_y(x, y;\ t) &= -\omega_2 e^{-\omega_2^2 t/2} \sin(\omega_2 y + \phi_2). \end{aligned} \tag{74}$$

**Preliminaries: Calculation of the second moment matrix.** In (15) the components of the multi-scale second moment matrix are defined as

$$\mu_{ij}(\cdot, \cdot;\ t, s) = g(\cdot, \cdot;\ s) * (L_{x_i}(\cdot, \cdot;\ t)\, L_{xj}(\cdot, \cdot;\ t)), \tag{75}$$

where $g(x, y;\ s) = g^{(x)}(x;\ s) g^{(y)}(y;\ s)$ denotes the two-dimensional Gaussian kernel, and $g^{(x)}$ and $g^{(y)}$ are the one-dimensional Gaussian kernels in the $x$- and $y$-directions respectively. It is straightforward to show that in this case $\mu_{ij}$ at the origin $(x, y) = (0, 0)$ are given by

$$\begin{aligned} \mu_{11}(x, y;\ t, s) &= \frac{\omega_1^2 e^{-\omega_1^2 t}}{2} (1 - e^{-2\omega_1^2 s} \cos 2\phi_1), \\ \mu_{22}(x, y;\ t, s) &= \frac{\omega_2^2 e^{-\omega_2^2 t}}{2} (1 - e^{-2\omega_2^2 s} \cos 2\phi_2), \\ \mu_{12}(x, y;\ t, s) &= \omega_1 \omega_2\, e^{-(\omega_1^2 + \omega_1^2)(t+s)/2} \sin \phi_1\ \sin \phi_2. \end{aligned} \tag{76}$$

The validity of (**??**) and (**??**) follow from

$$\mu_{11}(x, y;\ t, s) = \omega_1^2\, e^{-\omega_1^2 t} g^{(x)}(x;\ s) * g^{(x)}(x;\ s) * \sin^2(\omega_1 x + \phi_1), \tag{77}$$

(which is obtained from (74) and (75) as well as similar calculations as for deriving (73) from (71)) and (76) can be obtained in an analogous way from

$$\mu_{12}(x, y;\ t, s) = \omega_1 \omega_2\, e^{-(\omega_1^2 + \omega_2^2) t/2} (g^{(x)}(x;\ s) * \cos(\omega_1 x + \phi_1))\, (g^{(y)}(y;\ s) * \cos(\omega_2 y + \phi_2)).$$

**Approximation: Large integration scale.** Assume now, for simplicity, that the integration scale is large ($\omega_1^2 s >> 1$). Then, the sensitivity to the actual position (given by $\phi_1$ and $\phi_2$) will be small, and the second moment matrix can be approximated by

$$\mu = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{pmatrix} = \begin{pmatrix} \omega_1^2 e^{-\omega_1^2 t}/2 & 0 \\ 0 & \omega_2^2 e^{-\omega_2^2 t}/2 \end{pmatrix}. \tag{78}$$

Under the assumption of weak isotropy, the estimate of the foreshortening $\epsilon = \cos \sigma$ is given by the ratio between the eigenvalues of $\mu$:

$$\epsilon = \frac{\sqrt{\mu_{11}}}{\sqrt{\mu_{22}}} = \frac{\omega_1 e^{-\omega_1^2 t/2}}{\omega_2 e^{-\omega_2^2 t/2}} = \frac{\omega_1}{\omega_2} e^{(\omega_2^2 - \omega_1^2) t/2} \tag{79}$$

showing that in the case of rotationally symmetric Gaussian smoothing, the estimate is unbiased only if the value of the scale parameter is zero. For strictly positive values of the scale parameter, the slant angle is systematically overestimated, due to the fact that the higher-frequency component $\cos(\omega_2 y + \phi_2)$ is suppressed faster by the scale-space smoothing operation than the lower-frequency component $\cos(\omega_1 x + \phi_1)$.

**Shape adaptation.** Assume next that the shape of the smoothing kernel is adapted to the local image structure, such that the pattern (71) is smoothed with scale value $t_1$ along the $x$-direction and scale value $t_2$ along the $y$-direction. Then, the second moment matrix computed with shape-adapted smoothing is given by

$$\mu = \begin{pmatrix} \omega_1^2 e^{-\omega_1^2 t_1}/2 & 0 \\ 0 & \omega_2^2 e^{-\omega_2^2 t_2}/2 \end{pmatrix}, \tag{80}$$

and the foreshortening estimate is

$$\hat{\epsilon} = \frac{\omega_1}{\omega_2} e^{(\omega_2^2 t_2 - \omega_1^2 t_1)/2}. \tag{81}$$

Clearly, the foreshortening estimate will be correct if the scale parameters are adapted such that

$$\frac{t_2}{t_1} = \left(\frac{\omega_1}{\omega_2}\right)^2. \tag{82}$$

This shows that the desired solution is always a fixed point for shape adaptation.

**Convergence condition.** To study the convergence properties of the specific scheme in section 6.2, introduce $\omega_1 = \omega$ and $\omega_2 = \omega/\epsilon$. Assume (as in the algorithm) that the smallest amount of smoothing is fixed, $t_2 = t_0$, and that $t_1$ is determined from the foreshortening estimate $\hat{\epsilon}$ according to $t_1 = t_2/\hat{\epsilon}^2$. Then, each new estimate $\hat{\epsilon}_{k+1}$ as function of a previous estimate $\hat{\epsilon}_k$ ($k \in \mathbb{Z}_+$) is given by

$$\hat{\epsilon}_{k+1} = \epsilon \frac{e^{-\omega^2 t_0/2\hat{\epsilon}_k^2}}{e^{-\omega^2 t_0/2\epsilon^2}} \tag{83}$$

where $\epsilon = \omega_1/\omega_2$ denotes the true foreshortening. Some rewriting gives

$$\frac{\hat{\epsilon}}{\epsilon}_{k+1} = \exp\left(-\frac{\omega^2 t_0}{2\epsilon^2}\left(\frac{1}{(\hat{\epsilon}_k/\epsilon)^2} - 1\right)\right). \tag{84}$$

Consider next the function

$$h(x) = \exp\left(-\frac{\omega^2 t_0}{2\epsilon^2}\left(\frac{1}{x^2} - 1\right)\right). \tag{85}$$

The fixed point $\hat{\epsilon}/\epsilon = 1$ is convergent if and only if the derivative of $h$ satisfies

$$|h'(x)|_{x=1} = \alpha_c = \frac{\omega^2 t_0}{\epsilon^2} < 1. \tag{86}$$

**Interpretation.** This expression gives a sharp condition on when the fixed point is convergent. By and large it means that the iterative procedure converges provided that the initial slant estimate is sufficiently close to the true value and provided that the initial amount of smoothing is sufficiently small compared to the wavelength of the signal. The more detailed convergence properties depend on the slant angle; the smaller the slant angle is the larger is the domain of convergence.

Given that $t_0$ is selected using the automatic scale selection method described previously (based on maxima over scales in the normalized anisotropy $\tilde{Q}$), the interpretation of this result is that the method can be expected to diverge if the noise level is too high compared to the wavelength, and that the method can be expected to converge provided that the noise level is sufficiently small. These qualitative effects can be observed experimentally.

**Uniqueness.** A final remark may be useful concerning the issue of uniqueness. From (84) it is clear that the number of fixed points is given by the number of roots to the equation $h(x) = x$. Since

$$h'(x) = -\frac{\alpha_c}{x^3}\, e^{-\alpha_c(\frac{1}{x^2}-1)/2} < 0 \quad \text{if } x > 0 \tag{87}$$

and $\partial_x x = 1 > 0$, it is obvious that only one root can exist. Hence, besides the singular cases when either $t_1$ or $t_2$ tend to infinity, the fixed point given by (82) is always unique.

## Acknowledgments

This work was partially performed under the ESPRIT-BRA project InSight and the ESPRIT-NSF collaboration Diffusion. The support from the Swedish Research Council for Engineering Sciences, TFR, and the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged.

A first version of this manuscript was presented in (Lindeberg & Gårding 1994). After the completion of this work, the following extensions and closely related works have been presented. (Weickert 1995) applies a non-linear diffusion scheme to the enhancement of ridge structures in finger-print images, where the conductivities in different directions are controlled by second moment descriptors. (Almansa & Lindeberg 1996) perform corresponding ridge enhancement based on the shape adaptation ideas in this article.

Extensions of the scale selection scheme reviewed in section 3.2, based on a $\gamma$-normalized derivative concept, have been presented in (Lindeberg 1996*a*, Lindeberg 1996*b*). A main advantage of this extension is that it allows for immediate localization when detecting one-dimensional image features, such as edges and ridges.

## References

Almansa, A. & Lindeberg, T. (1996), Enhancement of fingerprint images by shape-adapted scale-space operators, *in* J. Sporring, M. Nielsen, L. Florack & P. Johansen, eds, 'Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory', Kluwer Academic Publishers, Copenhagen, Denmark. (To appear).

Aloimonos, Y. (1988), 'Shape from texture', *Biological Cybernetics* **58**, 345–360.

Alvarez, L., Guichard, F., Lions, P.-L. & Morel, J.-M. (1993), 'Axioms and fundamental equations of image processing', *Arch. for Rational Mechanics* **123**(3), 199–257.

Babaud, J., Witkin, A. P., Baudin, M. & Duda, R. O. (1986), 'Uniqueness of the Gaussian kernel for scale-space filtering', *IEEE Trans. Pattern Analysis and Machine Intell.* **8**(1), 26–33.

Bajcsy, R. & Lieberman, L. (1976), 'Texture gradients as a depth cue', *Computer Vision, Graphics, and Image Processing* **5**, 52–67.

Bergen, J. R., Anandan, P., Hanna, K. J. & Hingorani, R. (1992), Hierarchical model-based motion estimation, *in* G. Sandini, ed., 'Proc. 2nd European Conf. on Computer Vision', Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag, Santa Margherita Ligure, Italy, pp. 237–252.

Bigün, J. (1990), 'A structure feature for some image processing applications based on spiral functions', *Computer Vision, Graphics, and Image Processing* **51**, 166–194.

Bigün, J., Granlund, G. H. & Wiklund, J. (1991), 'Multidimensional orientation estimation with applications to texture analysis and optical flow', *IEEE Trans. Pattern Analysis and Machine Intell.* **13**(8), 775–790.

Blake, A. & Marinos, C. (1990*a*), 'Shape from texture: estimation, isotropy and moments', *J. of Artificial Intelligence* **45**, 323–380.

Blake, A. & Marinos, C. (1990*b*), Shape from texture: the homogeneity hypothesis, *in* 'Proc. 3rd Int. Conf. on Computer Vision', IEEE Computer Society Press, Osaka, Japan, pp. 350–353.

Blakemore, C. (1970), 'A new kind of stereoscopic vision', *Vision Research* **10**, 1181–1200.

Blostein, D. & Ahuja, N. (1989), 'Shape from texture: integrating texture element extraction and surface estimation', *IEEE Trans. Pattern Analysis and Machine Intell.* **11**(12), 1233–1251.

Brown, L. G. & Shvaytser, H. (1990), 'Surface orientation from projective foreshortening of isotropic texture autocorrelation', *IEEE Trans. Pattern Analysis and Machine Intell.* **12**(6), 584–588.

Cipolla, R., Okamoto, Y. & Kuno, Y. (1993), Robust structure from motion using motion parallax, *in* 'Proc. 4th Int. Conf. on Computer Vision', Berlin, Germany, pp. 374–382.

Florack, L. M. J., Salden, A. H., t. Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1995), 'Nonlinear scale-space', *Image and Vision Computing* **13**, 279–294.

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1992), 'Scale and the differential structure of images', *Image and Vision Computing* **10**(6), 376–388.

Förstner, M. A. & Gülch, E. (1987), A fast operator for detection and precise location of distinct points, corners and centers of circular features, *in* 'Proc. Intercommission Workshop of the Int. Soc. for Photogrammetry and Remote Sensing', Interlaken, Switzerland.

Gårding, J. (1992), 'Shape from texture for smooth curved surfaces in perspective projection', *J. of Mathematical Imaging and Vision* **2**, 329–352.

Gårding, J. (1993), 'Shape from texture and contour by weak isotropy', *J. of Artificial Intelligence* **64**(2), 243–297.

Gårding, J. & Lindeberg, T. (1994), Direct estimation of local surface shape in a fixating binocular vision system, *in* J.-O. Eklundh, ed., 'Proc. 3rd European Conference on Computer Vision', Vol. 800 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 365–376.

Gårding, J. & Lindeberg, T. (1996), 'Direct computation of shape cues using scale-adapted spatial derivative operators', *Int. J. of Computer Vision* **17**(2), 163–191.

Gibson, J. (1950), *The Perception of the Visual World*, Houghton Mifflin, Boston.

Jones, D. G. & Malik, J. (1992), Determining three-dimensional shape from orientation and spatial frequency disparities, *in* G. Sandini, ed., 'Proc. 2nd European Conf. on Computer Vision', Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag, Santa Margherita Ligure, Italy, pp. 661–669.

Kanatani, K. (1984), 'Detection of surface orientation and motion from texture by a stereological technique', *J. of Artificial Intelligence* **23**, 213–237.

Kanatani, K. & Chou, T. C. (1989), 'Shape from texture: general principle', *J. of Artificial Intelligence* **38**, 1–48.

Kimia, B. B., Tannenbaum, A. & Zucker, S. W. (1990), Toward a computational theory of shape: An overview, *in* 'Proc. 1st European Conf. on Computer Vision', Antibes, France, pp. 402–407.

Koenderink, J. J. (1984), 'The structure of images', *Biological Cybernetics* **50**, 363–370.

Koenderink, J. J. & van Doorn, A. J. (1976), 'Geometry of binocular vision and a model for stereopsis', *Biological Cybernetics* **21**, 29–35.

Koenderink, J. J. & van Doorn, A. J. (1990), 'Receptive field families', *Biological Cybernetics* **63**, 291–298.

Koenderink, J. J. & van Doorn, A. J. (1991), 'Affine structure from motion', *J. of the Optical Society of America* pp. 377–385.

Lindeberg, T. (1990), 'Scale-space for discrete signals', *IEEE Trans. Pattern Analysis and Machine Intell.* **12**(3), 234–254.

Lindeberg, T. (1993*a*), 'Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction', *J. of Mathematical Imaging and Vision* **3**(4), 349–376.

Lindeberg, T. (1993*b*), On scale selection for differential operators, *in* K. H. K. A. Høgdra, B. Braathen, ed., 'Proc. 8th Scandinavian Conf. on Image Analysis', Norwegian Society for Image Processing and Pattern Recognition, Tromsø, Norway, pp. 857–866.

Lindeberg, T. (1994*a*), On the axiomatic foundations of linear scale-space: Combining semi-group structure with causality vs. scale invariance, Technical Report ISRN KTH/NA/P--94/20--SE, Dept. of Numerical Analysis and Computing Science, KTH. Extended version to appear in J. Sporring and M. Nielsen and L. Florack and P. Johansen (eds.) Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory, Copenhagen, Denmark, Kluwer Academic Publishers, May 1996.

Lindeberg, T. (1994*b*), Scale selection for differential operators, Technical Report ISRN KTH/NA/P--94/03--SE, Dept. of Numerical Analysis and Computing Science, KTH. (Submitted).

Lindeberg, T. (1994*c*), *Scale-Space Theory in Computer Vision*, The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Dordrecht, Netherlands.

Lindeberg, T. (1996*a*), Edge detection and ridge detection with automatic scale selection, *in* 'Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition, 1996', San Francisco, California, pp. 465–470.

Lindeberg, T. (1996*b*), Feature detection with automatic scale selection, Technical Report ISRN KTH/NA/P--96/18--SE, Dept. of Numerical Analysis and Computing Science, KTH.

Lindeberg, T. & Gårding, J. (1993), Shape from texture from a multi-scale perspective, *in* H.-H. N. et. al., ed., 'Proc. 4th Int. Conf. on Computer Vision', IEEE Computer Society Press, Berlin, Germany, pp. 683–691.

Lindeberg, T. & Gårding, J. (1994), Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure, *in* J.-O. Eklundh, ed., 'Proc. 3rd European Conference on Computer Vision', Vol. 800 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 389–400.

Malik, J. & Rosenholtz, R. (1993), A differential method for computing local shape-from-texture for planar and curved surfaces, *in* 'Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition', pp. 267–273.

Manmatha, R. (1994), Measuring the affine transform using Gaussian filters, *in* J.-O. Eklundh, ed., 'Proc. 3rd European Conference on Computer Vision', Vol. 801 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 159–164.

Mumford, D. & Shah, J. (1985), Boundary detection by minimizing functionals, *in* 'Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition'.

Nitzberg, M. & Shiota, T. (1992), 'Non-linear image filtering with edge and corner enhancement', *IEEE Trans. Pattern Analysis and Machine Intell.* **14**(8), 826–833.

Nordström, N. (1990), 'Biased anisotropic diffusion: A unified regularization and diffusion approach to edge detection', *Image and Vision Computing* **8**, 318–327.

Osher, S. & Sethian, S. (1988), 'Fronts propagating with curvature dependent speed: algorithms based on the Hamilton-Jacobi formalism', *J. of Computational Physics* **79**, 12–49.

Pentland, A. P. (1986), 'Shading into texture', *J. of Artificial Intelligence* **29**, 147–170.

Perona, P. & Malik, J. (1990), 'Scale-space and edge detection using anisotropic diffusion', *IEEE Trans. Pattern Analysis and Machine Intell.* **12**(7), 629–639.

Rao, A. R. & Schunk, B. G. (1991), 'Computing oriented texture fields', *CVGIP: Graphical Models and Image Processing* **53**(2), 157–185.

Rogers, B. & Cagenello, R. (1989), 'Orientation and curvature disparities in the perception of three-dimensional surfaces', *Investigative Opthalmology and Visual Science* **30**, 262.

Saint-Marc, P., Chen, J.-S. & Medioni, G. (1991), 'Adaptive smoothing: A general tool for early vision', *IEEE Trans. Pattern Analysis and Machine Intell.* pp. 514–529.

Sapiro, G. & Tannenbaum, A. (1993), 'Affine invariant scale-space', *Int. J. of Computer Vision* **11**(1), 25–44.

Sato, J. & Cipolla, R. (1994), Extracting the affine transformation from texture moments, *in* J.-O. Eklundh, ed., 'Proc. 3rd European Conference on Computer Vision', Vol. 801 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 165–172.

Stevens, K. A. (1981), 'The information content of texture gradients', *Biological Cybernetics* **42**, 95–105.

Stone, J. V. (1990), Shape from texture: textural invariance and the problem of scale in perspective images of surfaces, *in* 'Proc. British Machine Vision Conference', Oxford, England, pp. 181–186.

Super, B. J. & Bovik, A. C. (1992), 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.

ter Haar Romeny, B., ed. (1994), *Geometry-Driven Diffusion in Computer Vision*, Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Dordrecht, Netherlands.

Terzopoulos, D. (1983), 'Multilevel computational processes for visual surface reconstruction', *Computer Vision, Graphics, and Image Processing* **24**, 52–95.

Tyler, C. W. & Sutter, E. E. (1979), 'Depth from spatial frequency difference: An old kind of stereopsis?', *Vision Research* **19**, 859–865.

Weber, J. & Malik, J. (1993), Robust computation of optical flow in a multi-scale differential framework, *in* 'Proc. 4th Int. Conf. on Computer Vision', Berlin, Germany, pp. 12–20.

Weickert, J. (1995), Multiscale texture enhancement, *in* '6th International Conference on Computer Analysis of Images and Patterns', Prague.

Wildes, R. P. (1981), 'Direct recovery of three-dimensional scene geometry from binocular stereo disparity', *IEEE Trans. Pattern Analysis and Machine Intell.* **13**(8), 761–774.

Witkin, A. P. (1981), 'Recovering surface shape and orientation from texture', *J. of Artificial Intelligence* **17**, 17–45.

Witkin, A. P. (1983), Scale-space filtering, *in* 'Proc. 8th Int. Joint Conf. Art. Intell.', Karlsruhe, West Germany, pp. 1019–1022.

Yuille, A. L. & Poggio, T. A. (1986), 'Scaling theorems for zero-crossings', *IEEE Trans. Pattern Analysis and Machine Intell.* **8**, 15–25.