

Shape-Adapted Smoothing in Estimation of 3-D Depth 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 (KTH), S-100 44 Stockholm, Sweden

Abstract. Rotationally symmetric operations in the image domain may give rise to shape distortions. This article describes a way of reducing this effect for a general class of methods for deriving 3-D shape cues from 2-D image data, which are based on the estimation of locally linearized distortion of brightness patterns. By extending the linear scale-space concept into an *affine scale-space* representation and performing *affine shape adaption* of the smoothing kernels, the accuracy of surface orientation estimates derived from texture and disparity cues can be improved by typically one order of magnitude. The reason for this is that the image descriptors, on which the methods are based, will be relative invariant under affine transformations, and the error will thus be confined to the higher-order terms in the locally linearized perspective mapping.

1 Introduction

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 one cannot expect to know in advance at what scales the relevant image structures manifest themselves. 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 *et al.* 1986; Yuille and Poggio 1986; Lindeberg 1990, 1994; Koenderink

* *Proc. 3rd European Conf. on Computer Vision, (Stockholm, Sweden), May 2-5, 1994. In: Springer-Verlag Lecture Notes in Computer Science, vol. 800, pp. 389-400.*

and van Doorn 1990; Florack *et al.* 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 due to the use of rotationally symmetric Gaussian kernels. In edge detection, smoothing across “object boundaries” can affect both the shape and the localization of edges. A corresponding problem arises when deriving three-dimensional shape cues from image data. In shape-from-texture, rotationally symmetric smoothing affects the 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.

Non-uniform smoothing methods. To reduce the problems of shape distortion in edge detection, (Perona and 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

$$\partial_t L = \frac{1}{2} \nabla^2 L \tag{1}$$

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 $\partial_t L = \frac{1}{2} \nabla^T (c(x; t) \nabla L)$ so as to favour intra-region smoothing to inter-region smoothing. This approach has then been furthered into different evolution schemes, e.g., (Nordström 1990; Nitzberg and Shiota 1992; Alvarez *et al.* 1993; Sapiro and Tannenbaum 1993; Florack *et al.* 1993).

Affine shape-adapted smoothing. Improvements relative to rotationally symmetric smoothing can also be obtained using linear methods. As has been argued by several authors, it can in certain situations be advantageous to use filters that correspond to different scale values in different directions; e.g., a large scale value along an edge, and a smaller scale value in the perpendicular direction.

The subject of this article is to develop how linear, or affine, shape adaption of the smoothing kernels can be used as an important mechanism for reducing the shape distortions when deriving three-dimensional shape cues from image data. The basic 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 component (the derivative). The application of this general idea to vision problems goes back to (at least) (Koenderink and van Doorn 1976), and has been explored in several shape-from-X problems, such as shape-from-texture (Gårding 1992; Lindeberg and Gårding

1993a; Malik and Rosenholtz 1993), shape-from-disparity (Wildes 1981; Jones and Malik 1992; Gårding and Lindeberg 1993a), and motion estimation (Koenenink and van Doorn 1991; Weber and Malik 1993; Cipolla *et al.* 1993).

The general advantage of shape-adaption in shape-from-texture has been pointed out by (Stone 1990). He 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 a general class of shape-from-X methods can be improved by extending the traditional *linear scale-space* representation to an *affine scale-space* representation based on Gaussian kernels with arbitrary (positively definite) covariance matrices. We shall be mainly concerned with the monocular case, the shape-from-texture problem. The underlying idea is, however, of much wider generality, as will be illustrated on shape estimation from disparity gradients.

Because of space constraints, the presentation is condensed and proofs omitted. For details and further background, see (Lindeberg and Gårding 1993b).

2 Basic idea for shape adaption

To motivate the need for shape adaption and to illustrate the basic idea behind the presented approach, consider first a non-uniform Gaussian blob

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

as a simple linearized model of the projection of a rotationally symmetric Gaussian blob (where l_1 and l_2 are characteristic lengths in the x - and y -coordinate directions and g is the one-dimensional Gaussian, $g(x; t) = (2\pi t)^{-1/2} \exp(-x^2/2t)$). The foreshortening, ϵ , and the slant angle, σ , are given by $\epsilon = \cos \sigma = l_2/l_1$, 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 $L(x, y; t) = g(x; l_1^2 + t) g(y; l_2^2 + t)$. Thus, under scale-space smoothing the estimate of foreshortening varies as

$$\hat{\epsilon}(t) = \sqrt{\frac{l_2^2 + t}{l_1^2 + t}}, \quad (3)$$

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 $t_{\hat{\delta}}$, be related by

$$t_{\hat{t}} = t_{\hat{\delta}} \cos^2 \hat{\sigma}. \quad (4)$$

If this estimate is correct, then the slant estimate will be *unaffected* by the non-uniform smoothing operation. To illustrate this property, assume that the tilt estimate is correct ($\hat{\theta} = \theta = \pi/2$) and convolve the signal with a non-uniform Gaussian kernel $g(x, y; t_{\hat{i}}, t_{\hat{j}}) = g(x; t_{\hat{i}})g(y; t_{\hat{j}})$, which gives $L(x, y; t) = g(x; l_1^2 + t_{\hat{i}})g(y; l_2^2 + t_{\hat{j}})$. Then, the new foreshortening estimate is

$$\hat{\epsilon} = \epsilon(\hat{\sigma}; t_{\hat{i}}, t_{\hat{j}}) = \sqrt{\frac{l_2^2 + t_{\hat{j}}}{l_1^2 + t_{\hat{i}}}} = |\cos \sigma| \sqrt{1 + \frac{t_{\hat{j}}}{l_1^2 + t_{\hat{i}}} \left(\frac{\cos^2 \hat{\sigma}}{\cos^2 \sigma} - 1 \right)}. \quad (5)$$

Clearly, $\hat{\epsilon} = \epsilon$ if $\hat{\sigma} = \sigma$. In practice, however, we cannot assume that true values of (σ, θ) are known, since this requires knowledge about the solution to the problem we are to solve. A more realistic formulation is 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). Then, when a hypothesis about a certain surface orientation $(\hat{\sigma}_0, \hat{\theta}_0)$ has been established, the estimates can be improved iteratively

$$\hat{\sigma}_{k+1} = \arccos \epsilon(\hat{\sigma}_k; t_{\hat{i}}, t_{\hat{j}}) = h(\hat{\sigma}_k). \quad (6)$$

From the derivative of this mapping,

$$|(\partial_{\hat{\sigma}} h)(\hat{\sigma})| = |(\partial_{\hat{\sigma}} \arccos \epsilon)(\hat{\sigma}; t_{\hat{i}} \cos^2 \hat{\sigma}, t_{\hat{j}})| = \{\text{let } \hat{\sigma} = \sigma\} = \frac{t_{\hat{j}}}{l_1^2 + t_{\hat{i}}} < 1, \quad (7)$$

it is clear that the true value of $\hat{\sigma}$ is a *convergent fixed point* for (6). Hence, for the pattern (2) the method is guaranteed to converge to the true solution, provided that the initial estimate is sufficiently close to the true value.

Here, no assumptions have been made about what actual method should be used for computing surface orientation from image data. The example describes essential effects of the smoothing operation, which will arise in any shape-from-X method that contains a smoothing module and interprets a non-uniform Gaussian blob as the projection of a rotationally symmetric one.

3 Shape-from-texture and disparity-gradients: Review

Computational studies of shape-from-texture and shape-from-stereo-cues have been done by several researchers. For a literature survey, see the previously cited papers on the subjects. Here, we shall consider the approach by (Lindeberg and Gårding 1993a; Gårding and Lindeberg 1993b), which will be briefly reviewed.

Measuring local affine distortion. Let $L: \mathbb{R}^2 \rightarrow \mathbb{R}$ be the image brightness, $\nabla L = (L_x, L_y)^T$ its gradient, and $w: \mathbb{R}^2 \rightarrow \mathbb{R}$ a window function. An image descriptor that allows for measurements of local linear distortions is *the windowed second moment matrix*. $\mu: \mathbb{R}^2 \rightarrow \text{SPSD}(2)$.² At any image point $q \in \mathbb{R}^2$ it is defined by

$$\mu_L(q) = \int_{x' \in \mathbb{R}^2} (\nabla L)(x') (\nabla L)^T(x') g(q - x') dx'. \quad (8)$$

² Here, $\text{SPSD}(2)$ stands for the cone of symmetric positive semidefinite 2×2 matrices.

Let $R: \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined by $L(\xi) = R(B\xi)$, where B is an invertible 2×2 matrix representing a linear transformation. Then, $\mu_L(q)$ transforms as

$$\mu_L(q) = B^T \mu_R(p) B, \quad (9)$$

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

Shape-from-texture and disparity gradients. Given two measurements of μ_L and μ_R , (9) can be used for recovering B (up to an arbitrary rotation). This gives a direct method for deriving surface orientation from monocular cues, by imposing specific assumptions on μ_R , e.g., that μ_R is a constant times the unit matrix, $\mu_R = cI$ (weak isotropy), or that $\det \mu_R$ is locally constant (constant area). Similarly, if two cameras fixate the same surface structure, a direct estimate of surface orientation can be obtained provided that the vergence angle is known.

Scale selection. Computation of μ_L requires selection of two scale parameters, a *local scale* for computing derivatives, and an *integration scale* describing the size of the window function. The determination of these scales follows the scale selection method proposed by (Lindeberg 1993, 1994) based on extrema over scales of combinations of *normalized derivatives*, $\partial_{\xi_i} = \sqrt{t} \partial_{x_i}$. Integration scales are set to a constant γ (typically $\gamma = 1, \sqrt{2}$, or 2) times the scales where

$$\det \mu_{L,norm} = t^2 \det \mu_L \quad \text{or} \quad \text{trace } \mathcal{H}_{norm} L = t(L_{xx} + L_{yy}) \quad (10)$$

assume local maxima over scales. Then, local scales are selected from minima over scales of the *normalized anisotropy*,

$$\tilde{Q} = \frac{\sqrt{\text{trace}^2 \mu_L - 4 \det \mu_L}}{\text{trace } \mu_L}. \quad (11)$$

4 Affine 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 scale-space representation* generated by convolution with non-uniform Gaussian kernels. Given a symmetric positive semi-definite (covariance) matrix, $\Sigma_t \in \text{SPSD}(2)$, the non-uniform Gaussian kernel in the two-dimensional case can be defined by

$$g(x; \Sigma_t) = \frac{1}{2\pi\sqrt{\det \Sigma_t}} e^{-x^T \Sigma_t^{-1} x/2} \quad \text{where } x \in \mathbb{R}^2. \quad (12)$$

Then, given any $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, the affine scale-space representation of f can be defined as the three-parameter family of functions $L: \mathbb{R}^2 \times \text{SPSD}(2) \rightarrow \mathbb{R}$

$$L(\cdot; \Sigma_t) = g(\cdot; \Sigma_t) * f(\cdot). \quad (13)$$

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

$$f_L(\xi) = f_R(B\xi), \quad (14)$$

and define the affine scale-space representations by

$$L(\cdot; \Sigma_L) = g(\cdot; \Sigma_L) * f_L(\cdot) \quad \text{and} \quad R(\cdot; \Sigma_R) = g(\cdot; \Sigma_R) * f_R(\cdot), \quad (15)$$

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

$$L(\xi; \Sigma_L) = R(\eta; \Sigma_R) \quad \text{where} \quad \Sigma_R = B\Sigma_L B^T. \quad (16)$$

Hence, for any Σ_L there exists a Σ_R such that the affine scale-space representations of f_L and f_R are equal. This property does not hold for the traditional linear scale-space representation based on the rotationally symmetric Gaussian.

Interpretation in terms of eigenvectors and eigenvalues. Let $t_b \geq t_t > 0$ be the eigenvalues of Σ_t^{-1} , and let \bar{b} and \bar{t} be the corresponding eigenvalues. Then, convolution with (12) corresponds to (separable) smoothing with scale value t_b along the \bar{b} -direction and scale value t_t along the \bar{t} -direction.

Diffusion equation interpretation of affine scale-space. Rewrite Σ_t as $\Sigma_t = t\Sigma_0$, where $t \in \mathbb{R}_+$ and $\det \Sigma_0 = 1$. Then, with $\nabla = (\partial_x, \partial_y)^T$, the non-uniform scale-space satisfies the transformed diffusion equation $\partial_t L = \frac{1}{2} \nabla^T \Sigma_0 \nabla L$.

Fourier transform and semi-group property. From the Fourier transform of g , $G(\omega; \Sigma_t) = e^{-\omega^T \Sigma_t \omega / 2}$, it follows that the non-uniform Gaussian satisfies the semi-group property $G(\omega; \Sigma_1) * G(\omega; \Sigma_2) = G(\omega; \Sigma_1 + \Sigma_2)$.

5 Texture descriptors defined from affine scale-space

Given an image $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ with affine scale-space representation $L: \mathbb{R}^2 \times \text{SPSD}(2) \rightarrow \mathbb{R}$, let us now define the second moment matrix based on non-uniform smoothing $\mu_L: \mathbb{R}^2 \times \text{SPSD}(2)^2 \rightarrow \text{SPSD}(2)$ by

$$\mu_L(\cdot; \Sigma_t, \Sigma_s) = g(\cdot; \Sigma_s) * ((\nabla L)(\cdot; \Sigma_t) (\nabla L)(\cdot; \Sigma_t)^T) \quad (17)$$

where Σ_s represents the covariance matrix corresponding to the integration scale, and Σ_t the covariance matrix corresponding to the local scale.

Transformation property under linear transformations. Under a linear transformation of the image coordinates $\eta = B\xi$, this descriptor transforms as

$$\mu_L(q; \Sigma_t, \Sigma_s) = B^T \mu_R(Bq; B\Sigma_t B^T, B\Sigma_s B^T) B. \quad (18)$$

Shape adaption: Invariance property of fixed points. Assume next that we can compute the second moment matrix using shape adaption such that the scale matrix is equal to the inverse of the second moment matrix. Then, this fixed point will be preserved under affine transformations, and the backprojected window function (see (9)) transforms in the same way as the affine transformation.

To verify this property, consider a signal f_L , and assume that μ_L has been computed with shape adaption at a certain point $q_L \in \mathbb{R}^2$ such that

$$\mu_L(q_L; \Sigma_{t,L}, \Sigma_{s,L}) = M_L, \quad (19)$$

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

$$\Sigma_{t,L} = t M_L^{-1}, \quad \Sigma_{s,L} = s M_L^{-1}, \quad (t, s \in \mathbb{R}_+). \quad (20)$$

Moreover, define a transformed intensity pattern f_R by $f_L(\xi) = f_R(B\xi)$. 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 adaption gives

$$\mu_R(q_R; \Sigma_{t,R}, \Sigma_{s,R}) = M_R \quad \text{where} \quad \Sigma_{t,R} = t M_R^{-1}, \quad \Sigma_{s,R} = s M_R^{-1}. \quad (21)$$

Similar properties hold for the scale-space maxima of $\det \mu_L$ and $\det \mathcal{H}_{norm} L$ as well, i.e., the entities used for selecting integration scales (and interest points).

Interpretation in the case of weak isotropy. At first glance, this property may be regarded as somewhat arbitrary. However, it has a simple geometric interpretation when the surface pattern is weakly isotropic. If the second moment matrix of the surface pattern is proportional to the unit matrix, then $M_R = cI$, and at the fixed point the covariance matrices $\Sigma_{t,R}$ and $\Sigma_{s,R}$ are also proportional to the unit matrix. This corresponds to rotationally symmetric smoothing and rotationally symmetric window functions in the tangent plane to the surface.

6 Designing an iterative procedure

Although the entity used for shape adaption is directly measurable from the second moment matrix, there is a chicken-and-the-egg aspect in the problem. The goal is to estimate the second moment matrix, while the smoothing procedure requires this information to obey the invariance properties. Nevertheless, an *iterative procedure* can be formulated. The invariance properties are obtained provided that the procedure converges to the desired fixed point.

Variation of scale matrices. A general variation of each scale matrix leads to a three-parameter variation. Hence, if no restrictions are imposed, there are six parameters to vary. The dimensionality of this search space can, however, be reduced by coupling the local and integration scale matrices such that

$$\Sigma_t = t \Sigma_0, \quad \Sigma_s = s \Sigma_0, \quad (22)$$

for some matrix Σ_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 Σ_0 , and the other two determining the size of each scale matrix). Here, we shall consider the following methods for choosing these parameters:

Shape of the smoothing matrices. With reference to the fixed point condition, the shape of Σ_0 will be selected *proportional to μ_L* at the given point.

Size of the integration scale matrix. To preserve the invariance of the scale-space maximum under linear transformations, one may 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 $\det \mathcal{H}_{norm}L$) *assumes a maximum over scales*. In practice, it is often reasonable to keep s constant.

Size of the local scale matrix. Further precautions must be taken when choosing the size of local scale matrix. Choosing the scale that maximizes the normalized anisotropy \tilde{Q} over scales is not appropriate in the non-isotropic case. To formulate an alternative criterion, one may at first consider letting $\det \Sigma_t$ be constant during the iterations. Such an approach would, however, lead to systematic overestimates of the anisotropy and, hence, the slant angle. A simple explanation for this is that the amount of smoothing in the direction of the smallest eigenvalue of the local scale matrix would be smaller in the non-isotropic case than in the isotropic case. Hence, the contribution from the fine scale variations in the signal (noise) would be expected to be larger and the anisotropy would increase.

Here, we propose to keep the *smallest eigenvalue* of Σ_0 constant during the iterations. By this approach the smallest amount of smoothing will be preserved, and the same (minimum) degree of noise suppression is guaranteed.

Composed method for shape adaption. To summarize, a straightforward method can be expressed as follows ($M_I^{(k)}$ denotes the matrix for shape adaption in the k th iteration step, $\mathcal{D}_{norm}L$ 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. $s_{\mathcal{D}L}^{(k)} = \begin{cases} \gamma^2 \cdot (s : \max_{s>0} (\mathcal{D}_{norm}L)(q; sM_I^{(k)})), & \text{or} \\ s_{\mathcal{D}L}^{(0)} & \text{when } k \geq 1. \end{cases}$
4. $M^{(k+1)} = \mu_L^{-1}(q; tM_I^{(k)}, s_{\mathcal{D}L}^{(k)}M_I^{(k)})$. Go to step 2 if not converged.

A useful convergence criterion is to require the angle between two successive surface normals (computed, e.g., under the assumption of weak isotropy) to be less than some given threshold $\Delta\varphi < \varepsilon$ (and to limit the number of iterations).

7 Experiments

This shape adaption scheme has been integrated with the shape-from-texture and shape-from-disparity-gradient methods reviewed in section 3. Experiments have been performed on real and synthetic reference data 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 1.)

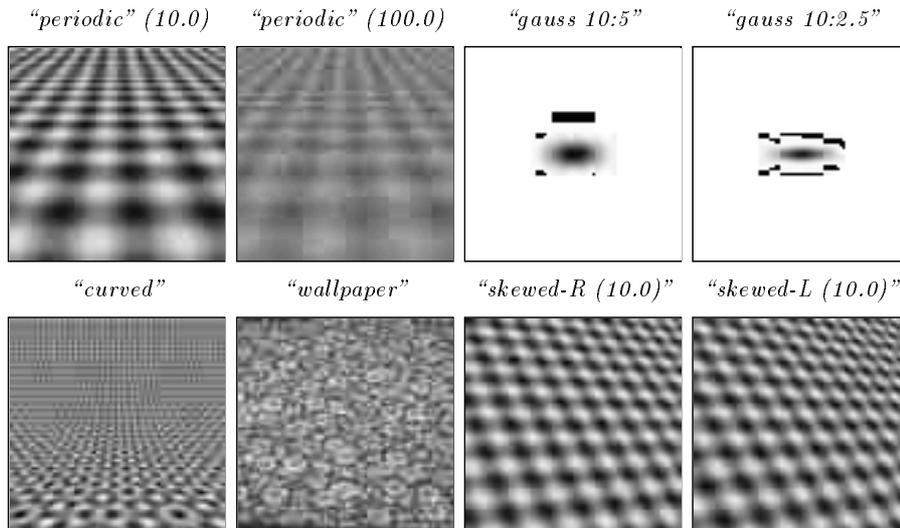


Fig. 1. Grey-level images used in the experiments on shape-from-texture. The first image, called 'periodic', is shown with added Gaussian noise of standard deviation 10.0 and 100.0, respectively. For the other test data only the original image is shown.

Shape-from-texture. Table 1 shows the result of applying the scheme to the image labelled "periodic" and using different amounts of added noise. To compare the effect of shape adaption 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*, γ .³ Notice that for high noise levels the shape adaption leads to improvements in estimated surface orientation that cannot be achieved by just increasing the integration scale. Moreover, substantial improvements can be obtained after just one or two iterations. Corresponding experimental results for the other images are given in tables 2–3.

Shape-from-disparity-gradients. The need for shape adaption 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 the basis for inferring cues to the three-dimensional surface structure.

Table 4 shows the results of including shape adaption as an essential step in the shape-from-disparity-gradient method indicated in section 3 and applying it to the stereo pair in figure 1. (The shape adaption has been performed on both images independently before the surface orientation estimates have been computed from (9).) Observe how the error decreases with the iterations.

³ This parameter determines how large the integration scale is relative to the scale at which the maximum in the scale-space signature is assumed (see section 3). Basically, a larger value of γ can be expected to increase the accuracy up to the point where the modelling error increases due to violations of the local linear approximation.

periodic (1.0) ($\gamma = 1.0$)		periodic (10.0) ($\gamma = 1.0$)		periodic (100.0) ($\gamma = 1.0$)	
0 (56.94, 80.94)	8.29	0 (56.31, 81.13)	8.37	0 (47.11, 74.45)	17.90
1 (60.03, 86.52)	3.00	1 (60.18, 86.32)	3.18	1 (54.18, 81.32)	9.31
2 (60.10, 87.55)	2.12	2 (60.42, 87.55)	2.16	2 (56.39, 84.21)	6.09
periodic (1.0) ($\gamma = 1.4$)		periodic (10.0) ($\gamma = 1.4$)		periodic (100.0) ($\gamma = 1.4$)	
0 (58.88, 89.22)	1.29	0 (57.88, 88.95)	2.29	0 (53.86, 89.61)	6.14
1 (60.08, 90.12)	0.13	1 (60.51, 90.09)	0.52	1 (59.90, 90.76)	0.67
2 (60.09, 90.11)	0.13	2 (60.75, 90.11)	0.76	2 (61.45, 90.90)	1.65
periodic (1.0) ($\gamma = 2.0$)		periodic (10.0) ($\gamma = 2.0$)		periodic (100.0) ($\gamma = 2.0$)	
0 (60.29, 90.18)	0.33	0 (58.80, 90.04)	1.19	0 (49.95, 91.24)	10.10
1 (60.59, 89.95)	0.59	1 (60.49, 89.90)	0.49	1 (56.72, 90.74)	3.34
2 (60.59, 89.95)	0.60	2 (60.66, 89.90)	0.66	2 (58.38, 90.40)	1.65

Table 1. Shape adapted smoothing applied to the image labelled “periodic” (using different values of the relative integration scale γ). The columns show from left to right, the iteration index, slant and tilt values computed under the weak isotropy assumption, and the angle between the estimated and the true surface normal (reference values (60.0, 90.0)). (The first value within parentheses is the standard deviation of the noise, to be related to the grey-level range [0, 255]. All angles are given in degrees.)

gauss 10:5 (1.0)		gauss 10:5 (10.0)		gauss 10:5 (100.0)	
0 (66.71, 90.04)	6.71	0 (65.55, 89.52)	5.56	0 (66.71, 92.13)	6.98
1 (58.83, 90.01)	1.16	1 (58.66, 89.81)	1.34	1 (62.13, 90.64)	2.20
2 (60.10, 90.02)	0.10	2 (59.54, 89.73)	0.50	2 (61.72, 90.68)	1.82
gauss 10:2.5 (3.1)		gauss 10:2.5 (10.0)		gauss 10:2.5 (31.6)	
0 (80.68, 89.93)	5.16	0 (80.24, 89.90)	4.72	0 (79.36, 90.28)	3.85
1 (75.99, 90.00)	0.47	1 (75.89, 89.95)	0.37	1 (76.91, 90.17)	1.39
2 (75.26, 90.00)	0.25	2 (75.25, 89.95)	0.27	2 (75.93, 90.03)	0.41

Table 2. Shape adaption 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). (Relative integration scale: $\gamma = 1.0$.)

skewed (10.0) ($\gamma = 1.4$)		curved (10.0) ($\gamma = 1.4$)		wallpaper (10.0) ($\gamma = 16$)	
0 (28.81, 20.91)	1.26	0 (59.57, 92.07)	4.89	0 (46.09, 85.56)	4.71
1 (29.71, 20.95)	0.55	1 (55.99, 90.61)	1.11	1 (51.86, 85.44)	1.06
2 (29.75, 20.95)	0.53	2 (56.22, 90.73)	1.36	2 (53.74, 85.34)	2.94

Table 3. Shape adaption applied to the images labelled “skewed”, “curved”, and “wallpaper”. Reference orientations: (30.0, 20.0), (55.0, 90.0), and (50.8, 85.3) respectively.

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
5 (55.04, 60.99)	0.80	5 (53.19, 60.58)	1.45	5 (54.33, 60.61)	0.46

Table 4. Shape adaption applied to the stereo pair in figure 1. (Vergence angle: $2\delta = 10.00$. Reference orientation (in a cyclopean coordinate system): (54.60, 60.16).)

8 Summary and discussion

We have described a methodology for reducing shape distortions when computing three-dimensional shape cues from image data using operators (derivatives) defined from the linear scale-space representation. 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. If shape adaption can be performed such that the second moment matrix computed at a certain point is equal to the matrix used for shape adaption at that point, then this fixed point will be preserved under affine transformations of the brightness pattern. In the specific case when the surface pattern is weakly isotropy, this corresponds to rotationally symmetric operators in the tangent plane to the surface.

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 $f(x, y) = \cos\omega_1x + \cos\omega_2y$ ($\omega_1 < \omega_2$), the desired fixed point is convergent if the minimum amount of smoothing t_0 satisfies $\omega_1^2 t_0 < (\omega_1/\omega_2)^2$. For many images, there are (at least) two more fixed points; one corresponding to the shape adaption matrix being singular, and one corresponding to an infinite amount of isotropic smoothing.

The suggested scheme has an interesting relationship to non-linear diffusion schemes. If applied at edge points, it leads to more smoothing along the edge than in the perpendicular direction. In this respect, the work constitutes a link between processing modules based on sparse edge data and dense filter outputs.

References

- L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel: Axioms and fundamental equations of image processing. *Arch. Rat. Mech.*, (to appear).
- J. Babaud, A.P. Witkin, M. Baudin, and R.O. Duda: Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE-PAMI*, 8(1):26–33, 1986.
- R. Bajcsy and L. Lieberman: Texture gradients as a depth cue. *CVGIP*, 5:52–67, 1976.

- D. Blostein and N. Ahuja: Shape from texture: integrating texture element extraction and surface estimation. *IEEE-PAMI*, 11(12):1233–1251, 1989.
- L.G. Brown and H. Shvaytser: Surface orientation from projective foreshortening of isotropic texture autocorrelation. *IEEE-PAMI*, 12(6):584–588, 1990.
- R. Cipolla, Y. Okamoto, and Y. Kuno: Robust structure from motion using motion parallax. *4th ICCV*, 374–382, 1993.
- L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever: Scale and the differential structure of images. *IVC*, 10(6):376–388, 1992.
- : Non-linear scale-space. (submitted), 1993.
- M.A. Förstner and E. Gülch: A fast operator for detection and precise location of distinct points, corners and centers of circular features. *ISPRS*, 1987.
- J. Gårding: Shape from texture for smooth curved surfaces in perspective projection. *J. Math. Im. Vis.*, 2:329–352, 1992.
- J. Gårding and T. Lindeberg: Direct computation of shape cues by multi-scale retinotopic processing *IJCV*, (to appear). TRITA-NA-P9304, Royal Inst. Tech., 1993a.
- : Direct estimation of local surface shape in a fixating binocular vision system. *3rd ECCV*, (Stockholm, Sweden), (these proceedings), 1993b.
- D.G. Jones and J. Malik: Determining three-dimensional shape from orientation and spatial frequency disparities. *2nd ECCV*, 661–669, 1992.
- J.J. Koenderink: The structure of images. *Biol. Cyb.*, 50:363–370, 1984.
- J.J. Koenderink and A.J. van Doorn: Geometry of binocular vision and a model for stereopsis. *Biol. Cyb.*, 21:29–35, 1976.
- : Receptive field families. *Biol. Cyb.*, 63:291–298, 1990.
- : Affine structure from motion. *J. Opt. Soc. Am.*, 377–385, 1991.
- T. Lindeberg: Scale-space for discrete signals. *IEEE-PAMI*, 12(3):234–254, 1990.
- : Scale selection for differential operators. *8th Scand. Conf. Im. An.*, 857–866, 1993b.
- : *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- T. Lindeberg and J. Gårding: Shape from texture from a multi-scale perspective. *4th ICCV*, 683–691, 1993a.
- : Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D brightness structure. TRITA-NA-P9335, Royal Inst. Tech., 1993b.
- J. Malik and R. Rosenholtz: A differential method for computing local shape-from-texture for planar and curved surfaces. *CVPR*, 267–273, 1993.
- M. Nitzberg and T. Shiota: Non-linear image filtering with edge and corner enhancement. *IEEE-PAMI*, 14(8):826–833, 1992.
- N. Nordström: Biased anisotropic diffusion: A unified regularization and diffusion approach to edge detection. *IVC*, 8:318–327, 1990.
- P. Perona and J. Malik: Scale-space and edge detection using anisotropic diffusion. *IEEE-PAMI*, 12(7):629–639, 1990.
- G. Sapiro and A. Tannenbaum: Affine invariant scale-space. *IJCV*, 11(1):25–44, 1993.
- J.V. Stone: Shape from texture: textural invariance and the problem of scale in perspective images of surfaces. *Brit. Machine Vision Conf*, pp. 181–186, 1990.
- J. Weber and J. Malik: Robust computation of optical flow in a multi-scale differential framework. *4th ICCV*, 12–20, 1993.
- R.P. Wildes: Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *IEEE-PAMI*, 13(8):761–774, 1981.
- A. Witkin: Recovering surface shape and orientation from texture. *AI*, 17:17–45, 1981.
- : Scale-space filtering. *8th IJCAI*, 1019–1022, 1983.
- A. Yuille and T. Poggio: Scaling theorems for zero-crossings. *IEEE-PAMI*, 8:15–25, 1986.