

# Surface Orientation and Curvature from Differential Texture Distortion

Jonas Gårding\*

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

## Abstract

*A unified differential geometric framework for estimation of local surface shape and orientation from projective texture distortion is proposed, based on a differential version of the texture stationarity assumption introduced by Malik and Rosenholtz. This framework allows the information content of the gradient of any texture descriptor defined in a local coordinate frame to be characterized in a very compact form. The analysis encompasses both full affine texture descriptors and the classical “texture gradients”. For estimation of local surface orientation and curvature from uncertain observations of affine texture distortion, the proposed framework allows the dimensionality of the search space to be reduced from five to one.*

## 1 Introduction

*Shape-from-texture* is the problem of determining the shape and orientation of a three-dimensional surface from a static monocular image of it. Qualitative and quantitative analysis of this problem originated with the work of Gibson (1950), who introduced the concept of a *texture gradient*, i.e., the systematic change due to the combined effects of imaging and surface geometry in the size and shape of projected texture elements.

In computer vision, the pioneering work of Bajcsy and Lieberman (1976) largely follows Gibson’s texture gradient paradigm. In later work, however, the approach has often been a different one; see e.g. (Witkin, 1981; Blostein and Ahuja, 1989; Blake and Marinos, 1990a, 1990b; Gårding, 1993). Here the problem is formulated in terms of estimating the parameters of a global or regional surface model (typically a plane of unknown orientation). With the addition of a probabilistic texture model, this allows shape-from-texture to be treated as a statistical inference problem. Many

elegant computational techniques have been proposed, but they generally suffer from rather restrictive assumptions about the surface texture (e.g. isotropy) and/or the surface shape (e.g. planarity).

The texture gradient approach is free of these limitations, and allows the shape-from-texture problem to be formulated in terms of *local* estimation of surface orientation (two parameters) and surface curvature (three parameters). It is useful to decompose this problem into two relatively independent components; first, the problem of *estimating* projective texture distortion (the “texture gradients”) in an image, and second, the problem of *interpreting* the projective texture distortion in terms of surface orientation and shape. This purely geometric problem was addressed for planar surfaces in (Stevens, 1981), and for curved surfaces in (Gårding, 1992).

A natural way of formalizing the notion of a texture gradient is to consider *simple distortion gradients*, i.e., the normalized gradients of scalar-valued functions representing area or characteristic dimensions of projected texture elements. However, an inherent limitation of the simple distortion gradients is that they do not contain enough information to estimate the full surface curvature (Gårding, 1992). Malik and Rosenholtz (1994a, 1994b) therefore proposed to represent texture distortion in a more general way by the set of affine transformations that best map an image texture patch onto neighbouring patches. Each simple distortion gradient encodes some particular aspect of these transformations; for example, the area gradient corresponds to the gradient of the determinant of the linear part. The main advantage of using the full affine transformation is that it allows (in principle) complete recovery of both local surface orientation and surface curvature, at the cost of having to establish a more precise description of the local texture distortion.

In order to estimate the affine texture distortion in the image, Malik and Rosenholtz introduced the assumption of *texture stationarity*, which means that the surface texture elements can have arbitrary properties

---

\*This work was partially performed under the ESPRIT-BRA project INSIGHT, and with support from the Swedish National Board for Industrial and Technical Development, NUTEK.  
Mail address: NADA, KTH, S-100 44 Stockholm, Sweden.  
Email: jonasg@bion.kth.se

and shapes as long as they are all (locally and statistically) equivalent. This is a reasonable assumption which is significantly less restrictive than e.g. isotropy, and it is sufficiently precise to allow a strict mathematical analysis of its implications. The main purpose of the present paper is to derive a unified differential geometric framework for analysis of estimation of surface orientation and curvature from the gradient of (nearly) arbitrary texture descriptors, based on a strictly differential version of the stationarity assumption proposed by Malik and Rosenholtz (1994a, 1994b). By applying this framework to the problem of estimating the local surface parameters from noisy observations of texture distortion, it is shown that the dimensionality of the search space can be reduced from five to one.

## 2 Review of the geometric framework

In this section the basic differential geometric framework proposed in (Gårding, 1992) and used e.g. in (Malik and Rosenholtz, 1994a, 1994b; Lindeberg and Gårding, 1993; Gårding and Lindeberg, 1995) will be briefly reviewed. The notation and terminology follow O’Neill (1966).

Consider the perspective mapping of a smooth surface  $S$  onto a unit viewsphere  $\Sigma$ , as shown in Figure 1. At any point  $p$  on  $\Sigma$  let  $(\mathbf{p}, \mathbf{t}, \mathbf{b})$  be a local orthonor-

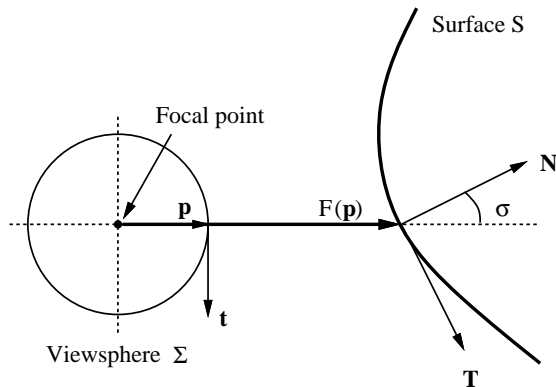


Figure 1: Local surface geometry and imaging model. The tangent planes to the viewsphere  $\Sigma$  at  $p$  and to the surface  $S$  at  $F(p)$  are seen edge-on but are indicated by the tangent vectors  $\mathbf{t}$  and  $\mathbf{T}$ . The tangent vectors  $\mathbf{b}$  and  $\mathbf{B}$  are not shown but are perpendicular to the plane of the drawing, into the drawing.

mal coordinate system with  $\mathbf{p}$  as view direction. The *tilt direction*  $\mathbf{t}$  is parallel to the direction of the gradient of the distance from the focal point, and  $\mathbf{b} = \mathbf{p} \times \mathbf{t}$ . Denote by  $F : \Sigma \rightarrow S$  the perspective backprojection, and by  $F_{*p} : T_p(\Sigma) \rightarrow T_{F(p)}(S)$  the derivative (linear approximation) of  $F$  at any point  $p$  on  $\Sigma$ , where  $T_p(\Sigma)$  is the tangent plane of  $\Sigma$  at  $p$ , and  $T_{F(p)}(S)$  is the tan-

gent plane of  $S$  at  $F(p)$ . In  $T_{F(p)}(S)$ , let  $\mathbf{T}$  and  $\mathbf{B}$  be the normalized images of  $\mathbf{t}$  and  $\mathbf{b}$  respectively. In the bases  $(\mathbf{t}, \mathbf{b})$  and  $(\mathbf{T}, \mathbf{B})$  the expression for  $F_{*p}$  is

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

where  $r = \|F(p)\|$  is the distance along the visual ray from the center of projection to the surface (measured in units of the focal length) and  $\sigma$  is the slant of the surface. The inverse eigenvalues of  $F_{*p}$ ,  $m < M$ , describe how a unit circle in  $T_{F(p)}(S)$  is transformed when mapped to  $T_p(\Sigma)$  by  $F_{*p}^{-1}$ ; it becomes an ellipse with  $m$  as minor axis (parallel to  $\mathbf{t}$ ) and  $M$  as major axis (parallel to  $\mathbf{b}$ ).

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

## 3 Local texture stationarity

The assumption of *texture stationarity* introduced by Malik and Rosenholtz (1994a, 1994b) provides a sound theoretical basis for analyzing all forms of shape-from-texture estimation based on differential texture distortion, using either full affine structure, texture gradients or any other texture property which is systematically affected by affine transformations. Intuitively, this assumption means that the surface texture elements can have arbitrary properties and shapes as long as there is no systematic geometric distortion among them.

We consider a texture element at some point  $P = F(p)$  in the surface to belong to the tangent plane of the surface at that point. In practice, this means that the extent of the texture element must be small relative to the radius of curvature of the surface at  $P$ . This local tangential texture model eliminates the theoretical difficulty involved in mapping a curved patch onto another patch with different Gaussian curvature, and is related in a natural way to the concept of differential texture distortion (e.g. texture gradients).

To formalize this texture model, we introduce a local orthonormal tangent frame field  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}}) \in T_{F(p)}(S)$ , which is stationary in some neighbourhood of  $P$  and for convenience is chosen such that  $\mathbf{T} = \tilde{\mathbf{T}}$  and  $\mathbf{B} = \tilde{\mathbf{B}}$  at  $P$ . Local texture stationarity can then be defined in terms of the frame field: *A texture is stationary in some region if each texture patch has the same representation (coordinates) with respect to  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$ .* As a consequence, the relation between the projective distortion of any texture feature and the local surface

orientation and curvature only depends on the projective distortion of the basis vectors  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$ ; there is no need to analyze each feature separately.

We must now define precisely what should be meant by (local) stationarity of the frame field  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$ . The property that the ‘‘shape’’ does not change is represented by the fact that  $\tilde{\mathbf{T}}$  and  $\tilde{\mathbf{B}}$  are orthogonal and have unit norm, so the only remaining issue is the orientation of the frame field. Intuitively, the frame field and thus the surface texture should remain parallel to itself when it is translated from  $P$  to some neighbouring point  $Q$ . This condition is satisfied if the frame field is moved by parallel transport along the geodesics originating at  $P$ , i.e., using a geodesic polar parameterization with pole  $P$  (O’Neill, 1966). For a planar surface this corresponds to the usual Euclidean notion of parallelism.

However, for a general curved surface the idea of a parallel texture field is, unfortunately, more problematic. For example, if a tangent vector  $\mathbf{v}$  is parallel-translated along some closed path  $\Gamma$ , the end result  $\mathbf{v}'$  will not necessarily be parallel to  $\mathbf{v}$ . (In fact, the *holonomy angle* between  $\mathbf{v}$  and  $\mathbf{v}'$  is equal to the total curvature of the region enclosed by  $\Gamma$ .) This inescapable fact of differential geometry means that, strictly speaking, the *orientation* of texture elements is a reliable feature only for surfaces with zero Gaussian curvature.

Nevertheless, the definition of parallelism in a neighbourhood of  $P$  based on a geodesic polar parameterization is perfectly valid; the problem is that for a surface with non-zero Gaussian curvature the result depends on  $P$ . For lack of a better alternative, however, we shall in the following apply the definition without restriction to zero Gaussian curvature, but the difficulties mentioned above should be borne in mind.

In fact, since the analysis is a differential one, the only practical consequence of the definition is that the covariant derivatives  $\nabla_{\mathbf{v}} \tilde{\mathbf{T}}$  and  $\nabla_{\mathbf{v}} \tilde{\mathbf{B}}$  have no components in  $T_{F(p)}(S)$ , for any tangent vector  $\mathbf{v}$ .

Finally, it is perhaps worth pointing out that the concept of texture stationarity is applicable to both geometric and statistical texture descriptors.

## 4 Differential texture distortion

Given the definition of local texture stationarity from the previous section, the effects of projection on a texture descriptor is fully defined by the effects of projection on the basis frame field  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$ . Let  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  denote the scaled images  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  of the frame vectors  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$ , defined by the property

$$\tilde{\mathbf{t}} = \frac{1}{m_p} F_*^{-1} \tilde{\mathbf{T}}, \quad \tilde{\mathbf{b}} = \frac{1}{M_p} F_*^{-1} \tilde{\mathbf{B}}.$$

$(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  represent the texture that is observed in the image. The constant scale factors  $1/m_p$  and  $1/M_p$  have been chosen to make  $\tilde{\mathbf{t}} = \mathbf{t}$  and  $\tilde{\mathbf{b}} = \mathbf{b}$  at  $p$ . Unlike  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$ , the vector field  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  is in general neither perpendicular nor of unit length away from the reference point; this distortion is what allows surface orientation and curvature to be estimated. The relation between the different vector fields that have been introduced is schematically depicted in Figure 2.

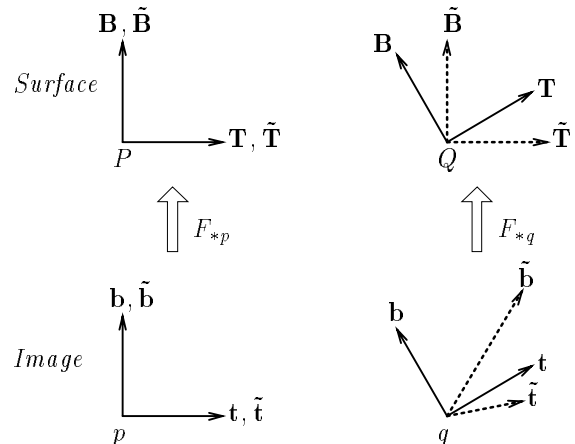


Figure 2: The frame fields  $(\mathbf{T}, \mathbf{B})$ ,  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$ ,  $(\mathbf{t}, \mathbf{b})$  and the vector field  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  shown at the reference point  $P = F(p)$  and another point  $Q = F(q)$ .

### 4.1 The fundamental equations

The definition of local texture stationarity in terms of the frame field  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$  allows a general analysis of a wide class of texture descriptors. In order to express the image gradient of a particular texture descriptor in terms of surface curvature and orientation, one only needs to know the rate of change (the covariant derivatives) of the projected frame field  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$ . These relations turn out to be simple and compact, and to emphasize their fundamental importance they are given as a proposition:

#### Proposition 1 (Locally stationary texture field)

Let  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$  be a locally stationary tangent frame field in a smooth surface  $S$ , which at  $P = F(p)$  coincides with  $(\mathbf{T}, \mathbf{B})$ . Let  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  be the images of  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$  scaled by constant factors to have unit length at  $p$ , i.e.,  $\tilde{\mathbf{t}} = (1/m_p) F_*^{-1} \tilde{\mathbf{T}}$  and  $\tilde{\mathbf{b}} = (1/M_p) F_*^{-1} \tilde{\mathbf{B}}$ , where  $m_p = \|F_*^{-1} \tilde{\mathbf{T}}(P)\|$  and  $M_p = \|F_*^{-1} \tilde{\mathbf{B}}(P)\|$ . The intrinsic covariant derivatives of  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  at  $p$  are given by

$$\nabla_{\tilde{\mathbf{t}}} \tilde{\mathbf{t}} = -(2 + r\kappa_t / \cos \sigma) \tan \sigma \mathbf{t}, \quad (2)$$

$$\nabla_{\tilde{\mathbf{b}}} \tilde{\mathbf{t}} = \nabla_{\tilde{\mathbf{t}}} \tilde{\mathbf{b}} = -r\tau \tan \sigma \mathbf{t} - \tan \sigma \mathbf{b}, \quad (3)$$

$$\nabla_{\tilde{\mathbf{b}}} \tilde{\mathbf{b}} = -r\kappa_b \sin \sigma \mathbf{t}. \quad (4)$$

A derivation can be found in (Gårding, 1995). In the following subsections Proposition 1 will be applied to analyze the information content of two different types of texture descriptors.

## 4.2 Full affine texture distortion

Malik and Rosenholtz (1994a, 1994b) estimated the affine transformation between an image patch and a number of nearby patches, and then used the parameters of the transformation to compute the local surface orientation and curvature. Starting from the differential framework described in Section 2, they derived an expression for the affine transformation which also involved some finite entities. A simpler relation is obtained directly from Proposition 1:

**Corollary 2 (Affine texture distortion)** *Let  $(\tilde{\mathbf{T}}, \tilde{\mathbf{B}})$  and  $(\tilde{\mathbf{t}}, \tilde{\mathbf{b}})$  be defined as in Proposition 1. Let  $\mathbf{w} = w_t \tilde{\mathbf{t}} + w_b \tilde{\mathbf{b}}$  (where  $w_t$  and  $w_b$  are constants) be the image of a locally stationary tangent vector field in  $S$ , and let  $\mathbf{v} = v_t \mathbf{t} + v_b \mathbf{b} \in T_p(\Sigma)$  be an arbitrary tangent vector at  $p$ . By definition, the first-order approximation of  $\mathbf{w}(p + \mathbf{v})$  is*

$$\mathbf{w}(p + \mathbf{v}) = \mathbf{w}(p) + \nabla_{\mathbf{v}} \mathbf{w},$$

which can be expressed as an affine transformation  $A_v$  of  $\mathbf{w}$ , parameterized by  $\mathbf{v}$ ,

$$\mathbf{w}(p + \mathbf{v}) = A_v \mathbf{w} = (I + M_v) \mathbf{w}, = (I + v_t M_t + v_b M_b) \mathbf{w},$$

where  $I$  is the identity, and

$$M_t = -\tan \sigma \begin{pmatrix} 2 + r\kappa_t / \cos \sigma & r\tau \\ 0 & 1 \end{pmatrix}, \quad (5)$$

$$M_b = -\tan \sigma \begin{pmatrix} r\tau & r\kappa_b \cos \sigma \\ 1 & 0 \end{pmatrix}, \quad (6)$$

in the  $(\mathbf{t}, \mathbf{b})$  basis.

The proof is a trivial application of Proposition 1; see (Gårding, 1995) for details.

To aid the interpretation of  $M_t$  and  $M_b$ , Figure 3 shows schematically the geometric effects of the components of a two-dimensional linear transformation.

Some examples of the local texture distortion patterns that result from (5) and (6) are shown in Figure 4. The drawings were generated by applying  $A_v$  to the central texture element<sup>1</sup> (a cross) in eight neighbouring positions, for various values of the surface curvature parameters and for two different step lengths  $(v_t, v_b)$ .

Corollary 2 reveals four invariant properties of the differential affine distortion patterns:

<sup>1</sup>To simplify the visual interpretation of these examples, the central texture element was chosen to be the orthographic projection of a cross. However, for the purposes of the analysis the shape of the element is irrelevant, as long as it is non-degenerate.

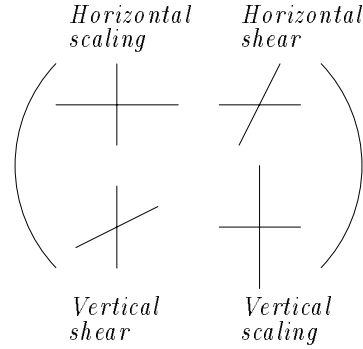


Figure 3: The geometric effect of the components of a 2-D linear transformation.

- $M_t[1, 2] = 0$ : There is no vertical shear in the  $\mathbf{t}$  direction.
- $M_b[2, 2] = 0$ : There is no vertical scaling in the  $\mathbf{b}$  direction. This observation is in keeping with the suggestion by Stevens (1981) that the tilt direction can be computed as the direction perpendicular to the direction of least variability, and the observation in (Gårding, 1992) that the *major gradient* vanishes in the  $\mathbf{b}$  direction independently of the surface curvature.
- $M_t[1, 2] = M_b[1, 1]$
- $M_t[2, 2] = M_b[2, 1]$

The first two properties have a simple geometric meaning which can easily be identified in Figure 4, whereas the third and fourth properties constrain the transformation in a more implicit way.

### Accuracy of the differential approximation.

The affine distortion map given by Corollary 2 can be interpreted as an approximation in two steps of the perspective projection of a locally stationary surface texture. First, the projective distortion in a small region is approximated by affine distortion; second, the *change* of this affine distortion in a neighbourhood is approximated by an affine function of image position. The accuracy of this approximation will in general depend on the precise surface shape, but it is nevertheless instructive to see what happens in the case of a very simple shape such as a plane. Figure 5 shows the image texture pattern generated according to Corollary 2 superimposed on a true perspective projection of the same surface pattern. The differences are almost impossible to see, so at least in this case the approximation errors are clearly not a cause for concern.

## 4.3 Texture gradients

Proposition 1 contains in a compact form all the information necessary to determine the information

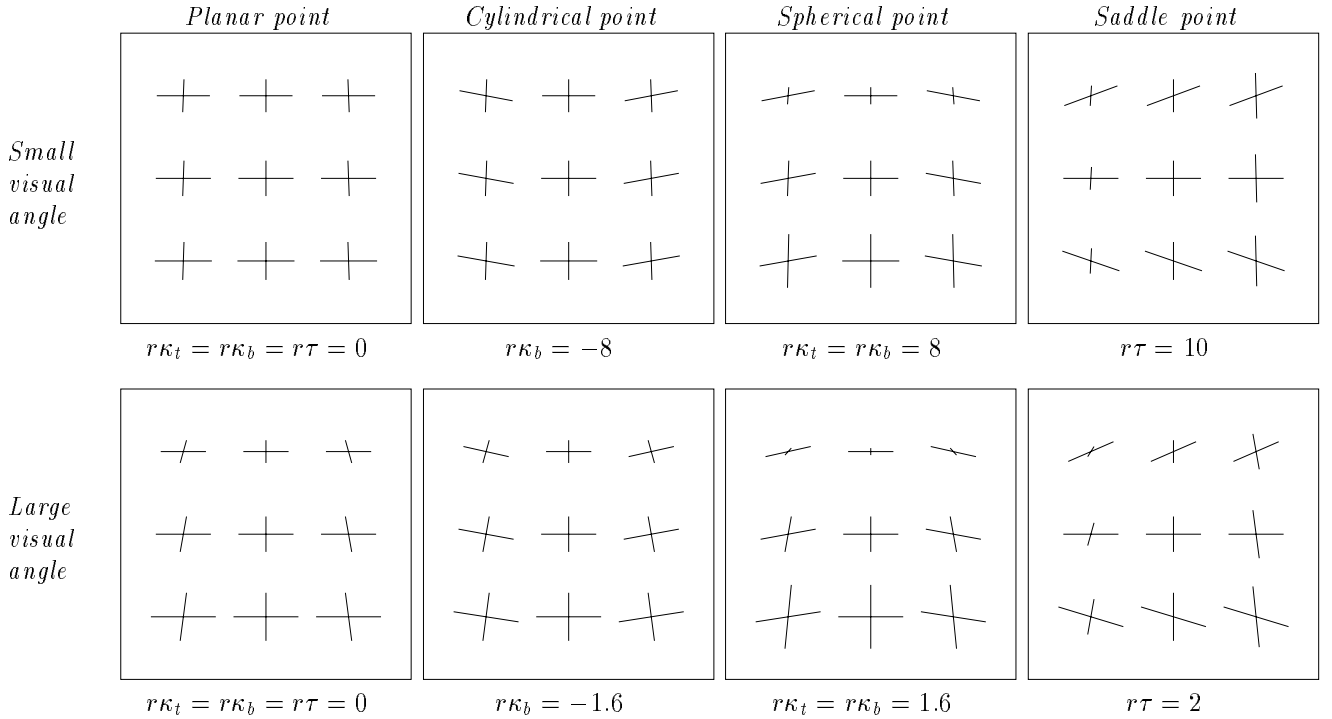


Figure 4: Examples of the effect of  $M_t$  and  $M_b$  at a planar point, a convex cylindrical point, a concave spherical point, and a saddle point. The surface orientation is  $\sigma = 50^\circ$ ,  $\theta = 90^\circ$  in all cases, which means that  $\mathbf{t}$  points towards the top and  $\mathbf{b}$  towards the left. The distance between the central texture element and its nearest neighbours is 0.03 focal units (corresponding to a visual angle of  $1.7^\circ$ ) in the top row, and 0.15 (visual angle  $8.5^\circ$ ) in the bottom row. The curvature parameters have been scaled accordingly to preserve the magnitude of their effects.

content of the classical “texture gradients” (Gibson, 1950; Stevens, 1981; Gårding, 1992), such as the minor and major gradients (sometimes referred to as compression and perspective gradients, respectively) or the area gradient.

As an example, let us derive an explicit expression for the minor gradient, which is the normalized gradient of a projected length  $m = k\|\tilde{\mathbf{t}}\|$  in the tilt direction (where  $k$  is an arbitrary scale factor). The normalized directional derivative at the point  $p$  in some direction  $\mathbf{v}$  can be expressed as

$$\frac{\mathbf{v}[m]}{m} = \frac{\mathbf{v}[k\|\tilde{\mathbf{t}}\|]}{k\|\tilde{\mathbf{t}}\|} = \mathbf{v}\left[\sqrt{\tilde{\mathbf{t}} \cdot \tilde{\mathbf{t}}}\right] = \tilde{\mathbf{t}} \cdot \nabla_{\mathbf{v}} \tilde{\mathbf{t}},$$

where we have used the fact that  $\|\tilde{\mathbf{t}}\| = 1$  at  $p$ . Hence, in the  $(\mathbf{t}, \mathbf{b})$  basis

$$\frac{\nabla m}{m} = \begin{pmatrix} \tilde{\mathbf{t}} \cdot \nabla_{\mathbf{t}} \tilde{\mathbf{t}} \\ \tilde{\mathbf{t}} \cdot \nabla_{\mathbf{b}} \tilde{\mathbf{t}} \end{pmatrix} = -\tan \sigma \begin{pmatrix} 2 + r\kappa_t / \cos \sigma \\ r\tau \end{pmatrix},$$

which agrees with the expression derived in (Gårding, 1992). Other texture gradients are computed analogously; for example, the normalized area gradient is given by the gradient of  $\|\tilde{\mathbf{t}} \times \tilde{\mathbf{b}}\|$ .

## 5 Shape-from-texture estimation

Up to this point the analysis has mainly concerned the “forward” case, i.e., characterization of various differential texture distortion measures in terms of surface orientation and curvature. The practical problem, however, is usually the inverse one: to estimate surface orientation and curvature from a number of observed local texture properties. Malik and Rosenholtz (1994a, 1994b) treated this problem for the case in which full affine descriptors can be computed (e.g. from spectrogram properties). Here it will be shown that this estimation problem can be simplified considerably, and at the same time generalized to incorporate an arbitrary covariance matrix for the measurement errors.

In theory, the affine distortion in two different directions uniquely determine the affine distortion in *any* direction, due to the linearity of covariant derivatives (see Corollary 2). In practice, however, estimates will be corrupted by noise, and it is therefore desirable to use measurements of the affine distortion in several directions. As pointed out by Malik and Rosenholtz (1994a, 1994b), a natural and theoretically well-founded estimation criterion is then to find the values

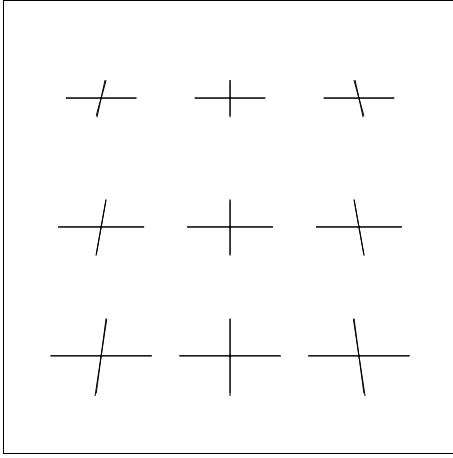


Figure 5: The affine approximation of projective distortion, superimposed on the corresponding true perspective image. The affine approximation image is copied from Figure 4 (bottom left). The images show identical and parallel crosses in a planar surface. The surface positions of the crosses are arbitrarily chosen by backprojecting a regular  $3 \times 3$  grid of image positions. The approximation errors (minute deviations of the vertical axes of some of the crosses) are almost impossible to see.

of  $(\sigma, r\kappa_t, r\kappa_b, r\tau, \theta)$  that minimize the sum of squared corrections to the measurements. Malik and Rosenholtz solved this five-dimensional minimization problem by a gradient descent technique. This problem is not intractable if a reasonably good starting approximation is available, but it is clearly computationally expensive. Starting from Corollary 2, a more direct method will be derived below.

### 5.1 The general case

To simplify the notation, we rewrite (5) and (6) as

$$M_t = \begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix}, \quad M_b = \begin{pmatrix} \beta & \delta \\ \gamma & 0 \end{pmatrix}.$$

To avoid dependence on the specific method used to estimate affine distortion, it is useful to (like Malik and Rosenholtz) consider an estimated affine distortion  $M = v_t M_t + v_b M_b$  in some direction  $(v_t, v_b)$  to be the basic observation. Without loss of generality, let the step be of unit length, i.e.,

$$v_t = \cos(\phi - \theta), \quad v_b = \sin(\phi - \theta),$$

where  $\phi$  is the angle between the direction in which  $M$  is measured and an arbitrary  $(x, y)$  coordinate frame, in which the (unknown) tilt direction is  $\theta$ .  $M(\phi)$  cannot be expressed in the  $(\mathbf{t}, \mathbf{b})$  basis since the orientation  $\theta$  of this frame is not yet known. By expressing it in the  $(x, y)$  frame instead, we obtain

$$M(\phi) = R_\theta (M_t \cos(\phi - \theta) + M_b \sin(\phi - \theta)) R_\theta^T, \quad (7)$$

where

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

This can be rewritten as

$$M(\phi) = M_x \cos \phi + M_y \sin \phi, \quad (8)$$

where

$$\begin{aligned} M_x &= R_\theta (M_t \cos \theta - M_b \sin \theta) R_\theta^T, \\ M_y &= R_\theta (M_t \sin \theta + M_b \cos \theta) R_\theta^T. \end{aligned}$$

Assuming that measurements have been made in  $n$  different directions  $\{\phi_i\}_{i=1}^n$ , let  $\mathbf{m}$  be a vector containing all the estimated transform components,

$$\mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_n \end{pmatrix}, \quad \text{where } \mathbf{m}_i = \begin{pmatrix} M(\phi_i)[1, 1] \\ M(\phi_i)[2, 1] \\ M(\phi_i)[1, 2] \\ M(\phi_i)[2, 2] \end{pmatrix}.$$

Analogously, let  $\hat{\mathbf{m}}(\alpha, \beta, \gamma, \delta, \theta)$  be the *predicted* components. The estimation problem can now be formulated as finding the parameter vector  $\mathbf{x} = (\alpha, \beta, \gamma, \delta, \theta)$  that minimizes the weighted corrections to the measurements, i.e.,

$$(\hat{\mathbf{m}}(\mathbf{x}) - \mathbf{m})^T C^{-1} (\hat{\mathbf{m}}(\mathbf{x}) - \mathbf{m}),$$

where  $C$  is the covariance matrix of the measurement errors.

The key observation that simplifies this problem considerably is that the prediction  $\hat{\mathbf{m}}(\mathbf{x})$  is *linear* in the first four parameters  $\mathbf{x}_4 = (\alpha, \beta, \gamma, \delta)$  of  $\mathbf{x}$ . By expanding (7), we obtain

$$\hat{\mathbf{m}} = J_\theta \mathbf{x}_4 = \begin{pmatrix} \cos \phi_1 J_x(\theta) + \sin \phi_1 J_y(\theta) \\ \vdots \\ \cos \phi_n J_x(\theta) + \sin \phi_n J_y(\theta) \end{pmatrix} \mathbf{x}_4,$$

where, with  $c = \cos \theta$  and  $s = \sin \theta$ ,

$$\begin{aligned} J_x(\theta) &= \begin{pmatrix} c^3 & -2c^2s & 2cs^2 & cs^2 \\ c^2s & -2cs^2 & -2c^2s & s^3 \\ c^2s & c(c^2 - s^2) & -s(c^2 - s^2) & -c^2s \\ cs^2 & s(c^2 - s^2) & c(c^2 - s^2) & -cs^2 \end{pmatrix}, \\ J_y(\theta) &= \begin{pmatrix} c^2s & c(c^2 - s^2) & -s(c^2 - s^2) & -c^2s \\ cs^2 & s(c^2 - s^2) & c(c^2 - s^2) & -cs^2 \\ cs^2 & 2c^2s & -2cs^2 & c^3 \\ s^3 & 2cs^2 & 2c^2s & c^2s \end{pmatrix}. \end{aligned}$$

Consequently, for any given tilt estimate  $\hat{\theta}$ , the optimal estimates of the remaining four parameters are obtained by solving the linear least squares problem

$$\min_{\mathbf{x}_4} (J_{\hat{\theta}} \mathbf{x}_4 - \mathbf{m})^T C^{-1} (J_{\hat{\theta}} \mathbf{x}_4 - \mathbf{m}).$$

As is well-known, the solution to this problem can be expressed in closed form as

$$\hat{\mathbf{x}}_4(\hat{\theta}) = (J_{\hat{\theta}}^T C^{-1} J_{\hat{\theta}})^{-1} J_{\hat{\theta}}^T C^{-1} \mathbf{m}. \quad (9)$$

This means that the original five-dimensional minimization problem in  $(\alpha, \beta, \gamma, \delta, \theta)$  has been reduced to a one-dimensional problem in  $\theta$ . For this problem even a crude strategy such as linear search in the interval  $(0, \pi)$  only results in a moderate computational cost.<sup>2</sup>

## 5.2 Further simplifications

If the directions  $\{\phi_i\}_{i=1}^n$  in which the affine distortion is estimated can be chosen freely, the problem can be simplified even further. In the experimental scheme used by Malik and Rosenholtz (1994a, 1994b), the affine distortion is estimated in  $n$  equally spaced directions in the interval  $(0, 2\pi)$ , and the measurements are assumed to be uncorrelated, i.e.,  $C = kI$ . (Without loss of generality, let  $k = 1$ .) It can then be shown that

$$(J_{\hat{\theta}}^T C^{-1} J_{\hat{\theta}})^{-1} = D = \frac{1}{n} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

This allows the optimal estimate  $\hat{\mathbf{x}}_4(\hat{\theta}) = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta})^T$  given by (9) to be expressed very concisely as

$$D \left( J_x^T(\theta) \sum_{i=1}^n (\mathbf{m}_i \cos \phi_i) + J_y^T(\theta) \sum_{i=1}^n (\mathbf{m}_i \sin \phi_i) \right),$$

The fact that the last two rows of  $J_x$  are equal to the first two rows of  $J_y$  can be used to speed up the evaluation of this expression. Moreover, it can be shown that the same result holds if the  $n$  directions are equally spaced in the interval  $(0, \pi)$  instead of in  $(0, 2\pi)$ .

If the directions  $\phi_i$  cannot be chosen in this way, a fast approximate solution can still be obtained by first solving a linear least squares problem to fit the form (8) to the data, and then treating the estimated matrices  $(M_x, M_y)$  as input data to the procedure described above. It turns out that the last column of  $M_x$  is equal to the first column of  $M_y$ , which means that only six (rather than eight) parameters are needed. However, there is clearly no guarantee that the result obtained by solving these two problems sequentially is close to the optimal solution to the original problem.

<sup>2</sup>The range of  $\theta$  is  $(0, 2\pi)$ , but since  $J(\theta + \pi) = -J(\theta)$  the value of the target function is the same for the estimates  $(\mathbf{x}_4, \theta)$  and  $(-\mathbf{x}_4, \theta + \pi)$ . This ambiguity is resolved by requiring  $\gamma = -\tan \sigma < 0$ .

## References

- Bajcsy, R. and Lieberman, L. (1976). Texture gradients as a depth cue. *Computer Graphics and Image Processing*, 5, 52–67.
- Blake, A. and Marinos, C. (1990a). Shape from texture: estimation, isotropy and moments. *J. of Artificial Intelligence*, 45, 323–380.
- Blake, A. and Marinos, C. (1990b). Shape from texture: the homogeneity hypothesis. In *Proc. 3rd Int. Conf. on Computer Vision*, pp. 350–353, Osaka, Japan. IEEE Computer Society Press.
- Blostein, D. and Ahuja, N. (1989). Shape from texture: integrating texture element extraction and surface estimation. *IEEE Trans. Pattern Anal. and Machine Intell.*, 11, no. 12, 1233–1251.
- 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, no. 2, 243–297.
- Gårding, J. (1995). Surface orientation and curvature from differential texture distortion. Technical Report TRITA-NA-P9510, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm.
- Gårding, J. and Lindeberg, T. (1995). Direct computation of shape cues based on scale-adapted spatial derivative operators. *Int. J. of Computer Vision*. (In press).
- Gibson, J. (1950). *The Perception of the Visual World*. Houghton Mifflin, Boston.
- Lindeberg, T. and Gårding, J. (1993). Shape from texture from a multi-scale perspective. In *Proc. 4th Int. Conf. on Computer Vision*, pp. 683–691, Berlin, Germany.
- Malik, J. and Rosenholtz, R. (1994a). Computing local surface orientation and shape from texture for curved surfaces. Technical Report UCB/CSD 93/775, Computer Science Division (EECS), University of California, Berkeley. (To appear in *Int. J. of Computer Vision*).
- Malik, J. and Rosenholtz, R. (1994b). Recovering surface curvature and orientation from texture distortion: a least squares algorithm and sensitivity analysis. In Eklundh, J.-O., editor, *Proc. 3rd European Conf. on Computer Vision*, volume 800 of *Lecture Notes in Computer Science*, pp. 353–364. Springer-Verlag.
- O’Neill, B. (1966). *Elementary Differential Geometry*. Academic Press, Orlando, Florida.
- Stevens, K.A. (1981). The information content of texture gradients. *Biological Cybernetics*, 42, 95–105.
- Witkin, A.P. (1981). Recovering surface shape and orientation from texture. *J. of Artificial Intelligence*, 17, 17–45.