PERGAMON

Pattern Recognition 35 (2002) 2141–2163

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

# Estimating the perspective pose of texture planes using spectral analysis on the unit sphere

Eraldo Ribeiro [1], Edwin R. Hancock *

*Department of Computer Science, University of York, York Y01 5DD, UK*

Received 29 December 2000; received in revised form 3 March 2001; accepted 29 May 2001

**Abstract**

Multiple vanishing point detection provides the key to recovering the perspective pose of textured planes. If vanishing points are to be detected from spectral information then there are two computational problems that need to be solved. Firstly, the search of the extended image plane is unbounded, and hence the location of vanishing points at or near infinity is difficult. Secondly, correspondences between local spectra need to be established so that vanishing points can be triangulated. In this paper, we offer a way of overcoming these two difficulties. We overcome the problem of unbounded search by mapping the information provided by local spectral moments onto the unit sphere. According to our representation, the position and direction of each local spectrum maps onto a great circle on the unit sphere. The need for correspondences is overcome by accumulating the great circle intercepts. Vanishing points occur at local accumulator maxima on the unit sphere. To improve the accuracy of the recovered perspective pose parameters for highly slanted planes, we use an adaptive spectral window. This selects the window so as to reduce spectral defocusing by minimising the determinant of the spectral covariance matrix. We experiment with the new shape-from-texture technique on both synthetic and real world data. Here it proves to be an accurate and robust means of estimating perspective pose.


*Keywords:* Shape-from-texture; Unit sphere; Evidence accumulation; Spectral analysis; Texture analysis; Planar surface recovery

## 1. Introduction

Shape-from-texture refers to the process of estimating of surface orientation from the perspective foreshortening of texture patterns [1,2]. In the literature, there are two routes to recovering the parameters of the required perspective projection. The first of these involves recovering the slant and tilt angles by spectral analysis. For instance both Krumm and Shafer [3], and Super and Bovik [4] have used affine approximations to recover the surface orientation through an iterative process of back-projection. However, this method is numerically cumbersome. The alternative approach is a geometric one. Here the idea is to use structural primitives to locate vanishing points [5,6]. In fact, Garding [7] has provided an elegant framework which relates the texture gradient to the underlying differential geometry of curved surfaces. Provided that two or more vanishing points are available, planar surface orientation can be directly determined. There are several limitations to the use of vanishing points. Firstly, there is a need to reliably segment structural texture primitives from the raw image data. Even if the primitives can be detected reliably, then

* Corresponding author. Tel.: +44-1904-43-3374; fax: +44-1904-43-2767.

*E-mail address:* erh@cs.york.ac.uk (E.R. Hancock).

[1] Supported by CAPES-BRAZIL, under grant: BEX1549/95-2.

0031-3203/02/$22.00 
PII: S0031-3203(01)00157-1

the search of the image plane for vanishing points is unbounded.

Recently, we have presented a new spectral analysis which combines some of the best features of the spectral and structural methods for perspective pose recovery from textured images. Our main theoretical contribution has been to show that local spectral peaks have a uniform angle along lines which radiate from a vanishing point [8,9]. The method avoids the use of time-consuming numerical back projection and the requirement for pre-segmented geometric texture tokens. However, we have demonstrated proof of concept for the method using very simple methods for locating and reconstructing vanishing points. One of the limitations of the methods used is the need to find correspondences between spectral peaks so that vanishing points can be triangulated. The aim of this paper is to explore ways of overcoming the dual problems of unbounded search for vanishing point and the need to acquire reliable spectral triangulation.

As revealed in the literature [10], the location of vanishing point from texture distribution is not itself a straightforward task. However, in order to overcome problems of poor feature detection many authors opt to use voting techniques such as the Hough-transform [11–13] and line-clustering [14]. Recently, Shufelt [10] has produced an empirical evaluation of some of the available methods. However, since we are working with local spectral peaks and are not concerned with line-primitive detection, one of the most elegant available methods is the unit-sphere representation of Barnard [15]. This maps the search for vanishing points to a Hough-like accumulation process on the unit sphere placed at the focal point on the optic axis. However, the method has been used only in the structural domain, where its effectiveness has been limited by the reliability of the available texture primitives. The aim of this paper is to explore the use of this search procedure in the spectral domain. The advantages of the method are twofold. Firstly, it bounds the search space for vanishing points to the unit sphere. Secondly, when used in conjunction with our property of spectral angle constancy, the accumulation process avoids the need for spectral correspondences for the purposes of triangulating vanishing points.

The idea behind our new method for perspective pose recovery is as follows. According to the representation of Barnard [15], lines on the image plane map to great circles on the unit sphere. Based on our spectral analysis, the position of a point on the image plane together with the direction of its spectral moment define a great circle on the unit sphere. When several local spectra originate from the same vanishing point, then they will have identical angles in the Fourier power spectrum. The Fourier domain angle of each detected spectral peak together with its position on the image plane define a straight-line on the image plane which maps to a great circle on the unit sphere. Corresponding spectral peaks which fall on the same line radiating from a vanishing point belong to the same great circle. Different spectral peaks at the same location are associated with different vanishing points. On the unit sphere, the locations of vanishing points correspond intersections of great circles. We treat the unit sphere as a Hough accumulator. Each detected spectral peak is converted into a great circle and we increment the number of votes in each cell traversed on the unit sphere. Vanishing points are located at the local maxima of accumulation.

To implement the method effectively, we need to accurately estimate the frequency-domain angles of the spectral peaks. Here there is a technical problem that must be overcome. If the window size used in the spectral estimator is mis-matched to the spectrum, then there will be problems of peak defocusing. This is particularly severe for highly inclined surfaces. Since our method relies on accurate peak localisation, this defocusing can lead to poor estimates of the spectral angles. We overcome this problem by employing a spectral window of adaptive size. We show how to set the size of the window so as to minimise the spectral dispersion.

To summarise, the novel contribution offered in this paper is to pose the problem of recovering the perspective pose of textured planes as the search for vanishing points on the unit sphere using Fourier domain spectral angles. This avoids the need for iterative optimisation or relaxation methods [4,3]. The method can be regarded as a frequency domain counterpart of several structural methods. For instance Kender's [5] agglomerative method uses the Hough-transform to recover vanishing points from line textures. Another structural approach to vanishing point recovery is to use the area gradient of the primitives. A simple version of this idea was reported by Ohta et al. [16] who places strong reliance on the ability to accurately estimate the primitive area. A more practical scheme has recently been reported by Kwon [6] who has shown how mathematical morphology can be used to determine the size of the texture primitives. However, this refinement is sensitive to the choice of the structuring element shape.

The outline of this paper is as follows. In Section 2, we review the perspective geometry of points on a plane. Section 3 extends this discussion to the frequency domain and shows how the power spectrum can be used to gauge perspective distortion. In Section 4, we review the key result underpinning the paper by showing that lines which radiate from vanishing points connect points of constant spectral angle. Sections 5 and 6, respectively, introduce the unit sphere representation and show how this representation can be combined with our constant spectral angle property to locate vanishing points. Section 7 presents a method for adaptively estimating the

window size used for spectral estimation. Section 8 provides experiments on both synthetic and real-world images. Finally, Section 9 summarises our conclusions and offers directions for future investigation.

## 2. Geometry modelling

### 2.1. Linear perspective projection

We commence by reviewing the projective geometry for the perspective transformation of points on a plane [17,18]. Specifically, we are interested in the perspective transformation between the object-centred co-ordinates of the points on the texture plane and the viewer-centred co-ordinates of the corresponding points on the image plane. To be more formal, suppose that the texture plane is a distance $h$ from the camera which has focal length $f < 0$. Consider two corresponding points. The point with co-ordinates $\mathbf{X_t} = (x_t, y_t, z_t)^{\mathrm{T}}$ lies on the texture plane while the corresponding point on the image plane has co-ordinates $\mathbf{X_i} = (x_i, y_i, x_i)^{\mathrm{T}}$. We represent the orientation of the viewed texture plane in the image plane co-ordinate system using the slant $\sigma$ and tilt $\tau$ angles. For a given plane, the slant is the angle between viewer line of sight and the normal vector of the plane. The tilt is the angle of rotation of the normal vector to the texture plane around the line of sight axis. Since we regard the texture as being "painted" on the texture plane, the texture height $z_t$ is always equal to zero. The local surface geometry is illustrated in Fig. 1.

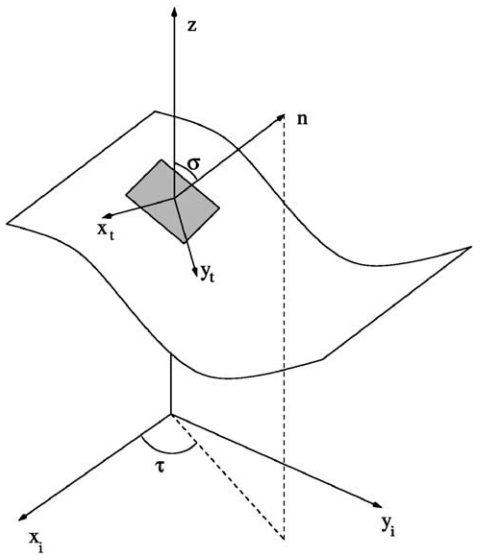



Fig. 1. Slant and tilt geometry of a local plane on the surface.

With these ingredients the perspective transformation between the texture-plane and image co-ordinate systems is given in matrix form by

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \frac{f}{h - x_t \sin\sigma} \times \left\{ \begin{bmatrix} \cos\sigma\cos\tau & -\sin\tau & \sin\sigma\cos\tau \\ \cos\sigma\sin\tau & \cos\tau & \sin\sigma\sin\tau \\ -\sin\sigma & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix} \right\}. \quad (1)$$

The first term inside the curly braces represents the rotation of the texture plane in slant and tilt. The second term represents the displacement of the rotated plane along the optic axis. Finally, the multiplying term outside the braces represents the non-linear foreshortening in the slant direction. When expressed in this way, $z_i$ is always equal to $f$ since the image is formed at the focal plane of the camera. As a result, we can confine our attention to the following simplified transformation of the $x$ and $y$ co-ordinates

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{f}{h - x_t \sin\sigma} \times \begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix} \begin{bmatrix} \cos\sigma & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix}. \quad (2)$$

This transformation can be represented using the shorthand $(x_i, y_i)^{\mathrm{T}} = T_p(x_t, y_t)^{\mathrm{T}}$, where $T_p$ is the $2 \times 2$ transformation matrix. As written above, the transformation $T_p$ can be considered as a composition of two transformations. The first of these is a non-uniform scaling proportional to the displacement in the slant direction. The second transformation is a counterclockwise rotation by an amount equal to the tilt angle.

The perspective transformation in Eq. (2) can be used to represent a non-linear geometric distortion of the surface texture pattern onto an image plane pattern. The methods presented in this paper use the spectral domain in order to recover the texture distortion in the image plane. Unfortunately, the non-linear nature of the perspective transformation makes the Fourier domain analysis of the texture somewhat intractable. To overcome this difficulty it is usual to use a linear or affine approximation of the perspective projection [18–21]. However, it is crucial that the linear approximation captures sufficient local information concerning the geometric distortion of the texture pattern to compute the parameters of surface orientation.

In order to proceed, we follow the bulk of the literature on shape-from-texture and derive a locally linear approximation to the perspective transformation [18–21]. The approximation we use can be considered as a variant of the paraperspective projection [16,19,21]. Paraperspective projections have the advantage of capturing the three main effects of the full perspective projection: foreshortening, viewing direction and distance effects. Therefore, the global quilting of the local approximations preserves the perspective effects required for recovering shape-from-texture. According to this linear approximation, the perspective distortion can be represented by the transformation matrix $T_*$. The paraperspective projection involves two steps. First, a point belonging to a given small region is orthographically projected along the direction of the line connecting the focal point of the camera to the centre of the region. This projection is done onto a hypothetical image plane parallel to the real image plane and passing through the centre of the region. Second, the point is then perspectively projected onto the real image plane. Paraperspective is a region to region projection. For a more complete discussion on linear alternatives to the perspective projection see Refs. [16,19,21].

This linear approximation can be found by performing a first-order Taylor expansion of the full perspective transformation matrix $T_p$ defined in Eq. (2). Suppose that the expansion point for the Taylor expansion on the image plane is $\mathbf{X_i^e} = (x_i^e, y_i^e)^{\mathrm{T}}$ on the image plane. We denote the local co-ordinate system relative to this expansion point on the image plane by $\mathbf{X}_i' = (x_i', y_i')^{\mathrm{T}}$ where $x_i = x_i' + x_i^e$ and $y_i = y_i' + y_i^e$. The first-order Taylor expansion is obtained by computing the Jacobian $J(\cdot)$ of $\mathbf{X_i}$ [22]. The required partial derivatives are calculated at the expansion point $\mathbf{X_i^e}$. After the necessary algebra, the resulting first-order Taylor approximation in terms of the slant and tilt angles is

$$T_*(\mathbf{X_i^e}) = \frac{\Omega}{hf\cos\sigma} \times \begin{bmatrix} x_i^e \sin\sigma + f\cos\tau\cos\sigma & -f\sin\tau \\ y_i^e \sin\sigma + f\sin\tau\cos\sigma & f\cos\tau \end{bmatrix}, \quad (3)$$

where $\Omega = f\cos\sigma + \sin\sigma(a_i\cos\tau + b_i\sin\tau)$. Hence, the linear transformation matrix $T_*$ depends only on the expansion point $\mathbf{X}_i^e = (x_i^e, y_i^e)^{\mathrm{T}}$, which is a constant, together with the slant and tilt angles, which are the goal of our analysis. This result is identical to the affine approximation of the perspective projection derived by Krumm and Shafer [17]. The only difference is that here we represent the linear approximation in terms of the slant and tilt angles rather than in terms of the normal vector slope parameters. A complete derivation of the transformation specified by the matrix given in Eq. (3) can be found in Ref. [17].

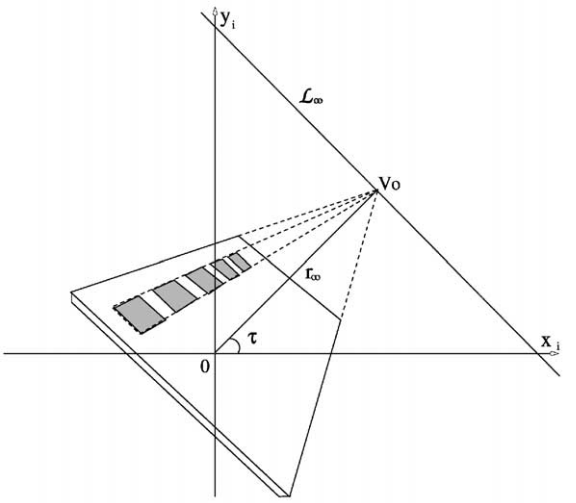



Fig. 2. Surface plane horizon line in a parametric normal form.

### 2.2. Vanishing points and the vanishing line

We now consider the effects of perspectivity on planar point-sets and the whereabouts of the associated vanishing points. The net effect of the global perspective transformation described in the previous section is to distort the viewer-centred texture pattern in the direction of vanishing points on the image plane [23,24]. Each vanishing point lies on a line called the vanishing line ($\mathscr{L}_\infty$) which is equivalent to the horizon line of the texture plane. Fig. 2 illustrates the geometry of the horizon or vanishing line. Following the geometry in Fig. 2, the horizon line can be represented in a parametric normal form by the equation

$$-f\frac{\cos\sigma}{\sin\sigma} = x_{vo}\cos\tau + y_{vo}\sin\tau, \quad (4)$$

where $r_\infty = -f\cos\sigma/\sin\sigma$ is the length of the normal from the line to the origin. In Eq. (4), the tilt angle $\tau$ is the tilt angle of the projected plane subtended between the normal to the horizon line and the $x_i$-axis. In the diagram, $\mathbf{V_o} = (x_{vo}, y_{vo})^{\mathrm{T}}$ is the horizontal vanishing point which is associated with a family of parallel lines in the tilt direction on the image plane [24]. Provided that the camera focal length $f$ is known, then Eq. (4) gives the equation of the vanishing line of the texture plane in terms of the slant and tilt angles. Using this representation of perspectivity, given two vanishing points $\mathbf{V_1} = (x_{v1}, y_{v1})^{\mathrm{T}}$ and $\mathbf{V_2} = (x_{v2}, y_{v2})^{\mathrm{T}}$ we can directly determine the horizon line $\mathscr{L}_\infty$ and hence the 3-D orientation of the plane in terms of the slant and tilt angles.

## 3. Projective distortion of the power spectrum

In this section we review the Fourier domain properties of affine transformations. These properties have been widely exploited in the literature on shape-from-texture [3,4,25]. Both Krumm and Shafer [3], and Super and Bovik [4] have used them in the recovery of planar pose using spectral texture information. It has also been exploited by Rosenholtz and Malik [25] in their work on local shape-from-texture.

The Fourier transform provides a representation of the spatial frequency distribution of a signal. We will commence by using a Fourier domain property of the affine transformation which has been well known for some time and which has recently been documented in the archival literature on the Fourier transform [26]. This property relates the linear effect of an affine transformation $A$ in the spatial domain to the frequency domain distribution. Suppose that $F(\cdot)$ represents the Fourier transform of the image. Furthermore, let $\mathbf{X}$ be a vector of spatial co-ordinates and let $\mathbf{U}$ be the corresponding vector of frequencies. According to Bracewell et al. [26], the distribution of image-plane frequencies $\mathbf{U_t}$ resulting from the Fourier transform of the affine transformation $\mathbf{X_i} = A\mathbf{X_t} + B$ is given by

$$F(\mathbf{U_i}) = \frac{1}{|det(A)|} e^{2\pi j \mathbf{U_t}^T A^{-1} \mathbf{B}} F[A^{-T}\mathbf{U_t}]. \quad (5)$$

In our case, the affine transformation is $T_A$ as given in Eq. (3) and there are no translation coefficients, i.e., $\mathbf{B} = \mathbf{0}$. As a result Eq. (5) simplifies to:

$$F(\mathbf{U_i}) = \frac{1}{|det(T_A)|} F[T_A^{-T}\mathbf{U_t}]. \quad (6)$$

In other words, the effect of the affine transformation of co-ordinates $T_A$ induces an affine transformation $T_A^{-T}$ on the texture-plane frequency distribution [3,18,27]. The spatial domain transformation matrix and the frequency domain transformation matrix are simply the inverse transpose on one-another.

We will consider here only the affine distortion of the frequency peaks, i.e., the energy amplitude will not be considered in the analysis. According to the affine approximation and Eq. (6), the way the Fourier domain transforms locally is governed by

$$\mathbf{U_i} = T_A^{-T}\mathbf{U_t}. \quad (7)$$

Fig. 3 illustrates the distortion of the spectral peaks. The left panel shows the transformation between the texture plane and the image plane. The right panel represents the corresponding distortion of a spectral peak taken in the local power spectrum. We refer to this angle as the spectral angle of the frequency peak $U_i$.

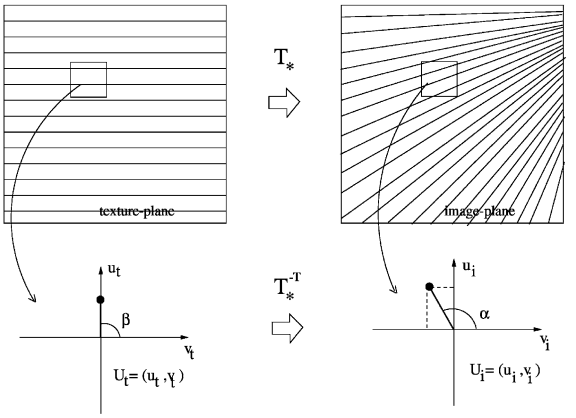



Fig. 3. Perspective distortion of the spectral angle. Parallel lines on the texture plane project to lines radiating from a vanishing point. In a local power spectrum, the spectral angle transforms analogously. The spectral angle is the average direction of all the directions present at that specific point.

## 4. Lines of constant spectral orientation

We have recently derived an interesting spectral property which simplifies the geometry involved in recovering perspective pose from spectral information [8]. The observation is a simple one. We consider spectral peaks in the power spectrum of the texture. When given a polar representation in the frequency domain, then each spectral peak can be characterised by the length of the radius vector from the origin and the angle between the $x$-axis and the radius vector. We have shown that for points which fall on a straight line which originates from a vanishing point, then the frequency domain angle remains constant [8,9]. Hitherto, we have used this property to show how vanishing points can be located by finding correspondences between points of identical spectral angle. Here we aim to use the property to accumulate votes on the unit sphere using the representation originally described by Barnard [15]. For completeness, in this section we review the derivation of the spectral property.

### 4.1. Spectral model of vanishing points

We are interested in understanding how the pattern of spectral peaks in the texture power spectrum is distorted by perspective geometry as we move across the image plane. In particular, we are interested in the relationship between the frequency domain polar angles of the spectral peaks and the positions of the vanishing points of the projected texture on the image plane. We simplify our analysis by confining our attention to a texture which consists of a family of parallel lines. This texture results in a spectrum which contains a peak with frequency vector $U_t = (u_t, v_t)^{\mathrm{T}}$. The frequency domain angle of this

spectral peak is $\beta = \arctan[v_t/u_t]$. The family of lines on the texture-plane lines has the parametric normal form:

$$\mathscr{L}_\beta: r_\beta = x_t \cos\beta + y_t \sin\beta. \tag{8}$$

We would like to compute the position of the image-plane vanishing point for this family of lines. To do this we substitute the equation of the family of texture plane lines into the perspective transformation of Eq. (2). By taking the limit of the resulting family of image-plane line as the texture plane co-ordinates tend to infinity, we can recover the position of the vanishing point on the image plane. To commence, we note that on the texture plane the equation governing the family of parallel lines is $y_t = (r_\beta - x_t \cos\beta)/\sin\beta$. Substituting this equation into Eq. (2) and taking the limit of $x_i$ and $y_i$ when $x_t$ tends to $\infty$ gives us the co-ordinates of the vanishing points in terms of the unprojected texture spectral angle:

$$x_v = \lim_{x_t\to\infty} f\frac{x_t \sin\beta\cos\tau\cos\sigma - \sin\tau(r - x_t\cos\beta)}{\sin\beta(h - x_t\sin\sigma)},$$

$$y_v = \lim_{x_t\to\infty} f\frac{x_t \sin\beta\sin\tau\cos\sigma + \cos\tau(r - x_t\cos\beta)}{\sin\beta(h - x_t\sin\sigma)} \tag{9}$$

After taking the limits, the vanishing point location is

$$(x_v, y_v)^{\mathrm{T}} = \frac{f}{\tan\beta\sin\sigma}(-\sin\tau + \tan\beta\cos\tau\cos\sigma, \cos\tau - \tan\beta\sin\tau\cos\sigma)^{\mathrm{T}}. \tag{10}$$

The above equation reveals that the location of the vanishing point is a function only of the direction of the original texture, the slant and tilt angles, and, the focal length of the camera.

### 4.2. Radiating lines and spectral peaks with constant angle

After the texture is projected onto the image plane under perspective geometry, the original texture spectral peak is transformed into the image-plane frequency vector $U_i = (u_i, v_i)^{\mathrm{T}}$. The geometrical meaning is that the family of parallel lines $\mathscr{L}_\beta$ which constitute the original texture (Eq. (8)) will transform to lines which intercept at a vanishing point, i.e. lines radiating from a vanishing point. We will call this family of lines $\mathscr{L}_\theta$. Additionally, the spectral peaks calculated at each region on the image plane reflect the directional tendency towards the vanishing point. The polar-angle associated with this image-plane frequency vector is

$$\alpha = \arctan\left[\frac{v_i}{u_i}\right]. \tag{11}$$

Using, the affine approximation to the perspective projection of frequency vectors given in Eq. (7), the spectral angle is given in terms of the components of the texture-plane frequency vector as

$$\alpha = \arctan\left[\frac{f\sin\tau + \tan\beta(x_i\sin\sigma + f\cos\tau\cos\sigma)}{f\cos\tau - \tan\beta(y_i\sin\sigma + f\sin\tau\cos\sigma)}\right]. \tag{12}$$

Next we consider the line connecting the point $(x_i, y_i)$ and the vanishing point $(x_v, y_v)$. The slope angle of this line is $\tan\theta = -(x_i - x_v)/(y_i - y_v)$ and the line has equation

$$\mathscr{L}_\theta: r_\theta = x_i\cos\theta + y_i\sin\theta. \tag{13}$$

Using the position of the vanishing point given in Eq. (10), the equation of the line connecting the two points is

$$x_i\sin\sigma\tan\theta + \tan^2\theta(y_i\sin\sigma + f\sin\tau\cos\sigma) + f[\tan\theta(\cos\tau\cos\sigma - \cos\tau) + \sin\tau] = 0. \tag{14}$$

Substituting for $x_i$ from the above equation into the expression for the projected polar angle of the spectral-peak from Eq. (12), it is straightforward to show that $\alpha = \theta$, $\forall(x_i, y_i) \in \mathscr{L}_\theta$. As a result, each line belonging to the family $\mathscr{L}_\theta$ connects points on the image plane whose the local spectral distributions have the same Fourier domain spectral angle $\alpha$. These lines will intercept at a unique point which is a vanishing point on the image plane. It is also important to stress that the system of equations is singular when the point $(x_i, y_i)$ is co-incident with the vanishing point. These observations are pivotal to the development of our new method for recovering perspective pose.

To provide a concrete illustration, Fig. 4 shows some examples of the lines of constant Fourier domain spectral angle computed using the perspective model developed in this section. In the figure we show examples of the lines of uniform spectral angle for a sinusoidal texture with frequency $\mathbf{U_t} = (0, 0.125)$ at several slant angles. For simplicity, we illustrate for tilt angle equal to zero. Fig. 5 shows some local power spectra sampled along

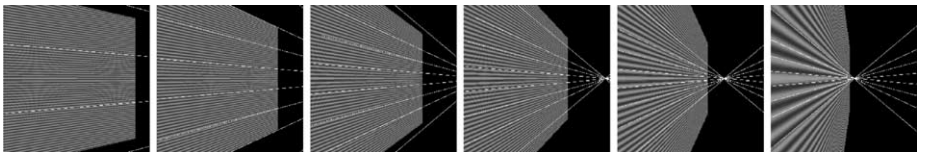


Fig. 4. Lines with same spectral angle together with the sinusoid image for different slant angles.

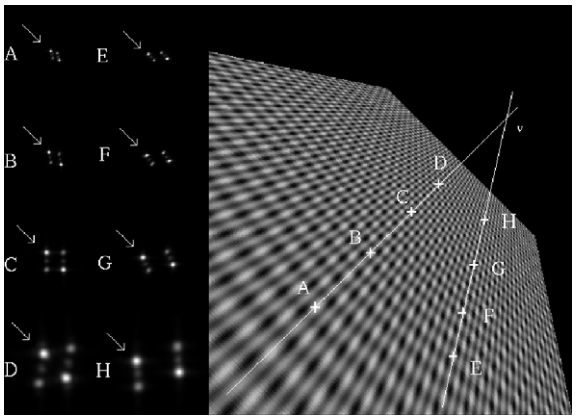



Fig. 5. Angular spectral components along lines radiating from a vanishing point.

two different lines that radiate from the vanishing point labelled $V$. The spectra are labelled by capital letters and are shown in the left-hand panel of the figure. In each case, the dominant peak frequency is marked by a white arrow. We have two groups of spectra belonging to two different straight lines in the figure. The first group is labelled with the letters A–D while the second group is labelled with the letters E–H. The main feature to note from these examples is that while the dominant frequency peak appears at the same angle within each group, it appears at different angles in the different groups.

Finally, it is important to stress that our method is only limited by the need for spectra with distinct and sharply focused peaks. As we will demonstrate later, this lends it to a diversity of real world textures.

## 5. Local spectral frequency under unit sphere mapping

In this section, we follow Barnard [15] and model the image plane in terms of spherical co-ordinates by projecting it onto a unit sphere centred at the optical centre. This projection simplifies the representation of the perspectivity of the texture plane and the search for its vanishing points. The main advantage is that unlike the image plane, the unit sphere is a closed space parametrised by the two angles of azimuth, and, zenith or elevation. Spherical projections of the image plane have been exploited by several authors [15,28]. However, they have employed structural representations of texture. Instead, we use the local spectral frequency to model texture.

The unit sphere is placed at the focal point and the image plane lies at a distance $f$ along the optical axis. Fig. 6 illustrates the projection geometry. As previously shown in Section 4, for each point on the image plane, the position $(x_i, y_i)$ and the measured frequency domain spectral angle $\alpha$ specify the equation of a line $\mathscr{L}_\alpha$ radiating from a vanishing point. We can write the equation of $\mathscr{L}_\alpha$ as $r_\alpha = x_i \cos\alpha + y_i \sin\alpha$. Each such line, $L_\alpha$ projects onto a great circle on the unit sphere. The great circles are constructed by intersecting the sphere by the plane that contains both the line $\mathscr{L}_\alpha$ and the centre of the sphere. Let $\mathbf{G}_\alpha$ be a radius vector that points from the centre of the unit sphere to a point on the corresponding great circle. If the point has azimuth angle $\varphi$ and elevation angle $\psi$, then the vector is given by

$$\mathbf{G}_\alpha = (\sin\varphi\cos\psi,\ \sin\psi,\ \cos\varphi\cos\psi)^{\mathrm{T}}. \tag{15}$$

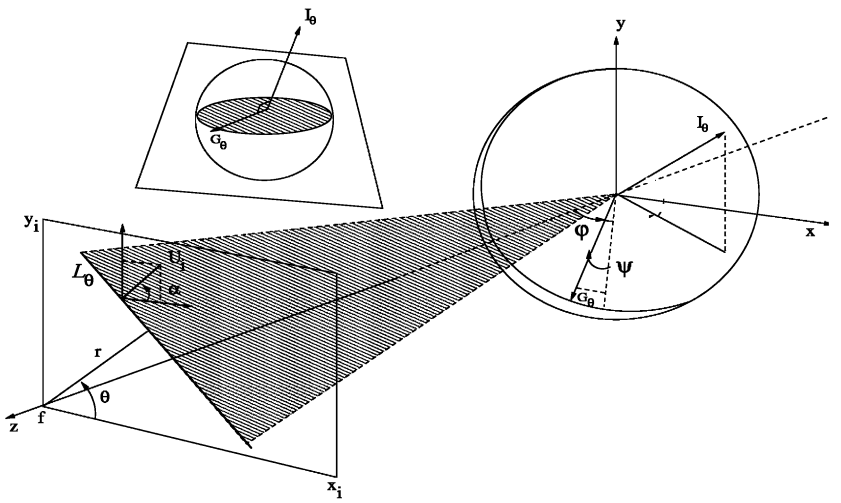



Fig. 6. Projecting lines from the image plane onto the unit Sphere.

Suppose that $I_\alpha$ is the normal vector to the plane which contains the great circle (shaded plane region in Fig. 6). This plane is called *interpretation plane* [15]. The normal vector of the interpretation plane can be obtained by choosing two points $P_1$ and $P_2$ on the line $\mathscr{L}_\alpha$. If $p_1$ and $p_2$ are the vectors from the origin of unit sphere to the points $P_1$ and $P_2$, then

$$I_\alpha = \frac{p_1 \times p_2}{\|p_1\| \, \|p_2\|}. \tag{16}$$

We choose the two points to be the intersections of the line $\mathscr{L}_\alpha$ with the image-plane co-ordinate axis. As a result, $p_1 = (0, r_\alpha/\sin\alpha, f)$ and $p_2 = (r_\alpha/\cos\alpha, 0, f)$. Fig. 6 shows in the geometry of the plane normal vector $I_\alpha$ and the great circle generator vector $\mathbf{G}_\alpha$. Since the vectors $\mathbf{G}_\alpha$ and $\mathbf{I}_\alpha$ are perpendicular to one another, they satisfy the condition $\mathbf{G}_\alpha.\mathbf{I}_\alpha = 0$. As a result, the azimuth and elevation angles of points on the great circle are related to the normal distance parameters ($r_\alpha$ and $\alpha$) of the straight line in the following manner

$$f(\sin\alpha \tan\psi + \cos\alpha \sin\varphi) = r_\alpha \cos\varphi. \tag{17}$$

## 6. Accumulation on the unit sphere

Having established the relationship between spectral angle and vanishing point location, we are now in a position to develop an accumulation algorithm on the unit sphere. We exploit the following two properties to map the search for vanishing points on to the unit sphere:

**Property 1** (Spectral frequency angle constancy). If $\mathscr{L}_\theta$ is a line radiating from a vanishing point on the image plane, then every local spectral distribution taken at points belonging to $\mathscr{L}_\theta$ will have a constant spectral angle $\alpha$. Conversely, each spectral angle $\alpha$ estimated from a local frequency distribution on the image plane specifies the equation of a line $\mathscr{L}_\theta$ which radiates from a corresponding vanishing point.

We can now exploit Property 1 to directly relate the local spectral angle to the equations of the great circles on the unit sphere. Using the equality between the angles $\theta$ and $\alpha$ and using the expression for a great circle in Eq. (17), we find

$$\psi = -\arctan\frac{f\cos\alpha\sin\varphi - r_\theta\cos\varphi}{f\sin\alpha}. \tag{18}$$

**Property 2** (From spectral frequency angles to great circles). Each spectral angle $\alpha$ estimated from the local frequency distribution centred at a point on the image plane maps to a great circle on the unit sphere. When several great circles intercept on the unit sphere, then the corresponding image-plane spectra will have originated from a common vanishing point.

This property is a direct corollary of Property 1. With this relationship to hand, we can now approach the problem of finding multiple vanishing points in a more robust way. Property 2 allows us to solve the two main problems of correspondence and unbounded search space. To proceed, in the following subsections we describe the technical details of our new method. The procedure for obtaining the slant and tilt angles from vanishing points is as follows:

- compute the local power spectra at a number of locations on the image plane.
- each spectral peak at each image location is transformed to a great circle on the unit sphere by computing its interpretation plane.
- the great circle intercepts on the unit sphere are computed using the Hough accumulation method outlined in the next section of this paper.
- the locations of the two most significant vanishing points are found by searching for accumulation maxima on the unit sphere.
- the vanishing points are back projected onto the image plane.
- the vanishing line is reconstructed and the slant and tilt parameters are estimated using Eq. (4).

### 6.1. Sampling the angles of the spectral peaks

To compute the spectral angle distribution, we require a way of sampling the local power spectrum. In particular we need a sampling procedure which provides a means of recovering the angle of the peaks of the power spectrum. We accomplish this by simply searching for local maxima over a filtered representation of the local power spectrum. We are interested in the dominant angular information rather than the entire frequency content of the power spectrum. As a result, we can ignore the very low frequency components of the power spectrum since these mainly describe micro-texture patterns or very slow energy variations [29]. We also filter the very low energy content of the spectrum. The very low energy component corresponds to weak line directions in the image. We are interested in only the most representative directions in the image. Another option for sampling the spectral peaks would be to use a bank of Gabor filters [18,29]. However, this procedure is time-consuming and would require a bank of filters with very fine angular coverage. In practice, this simple and direct search for local maxima throughout the image works well. We can use as many distinct spectral peaks as we can locate. However, a two-component decomposition is sufficient for our purposes.

We extract angular decompositions for the local power spectra at several locations on the image plane. The image sampling is done over a rectangular region over the portion of the image plane containing the texture. Since

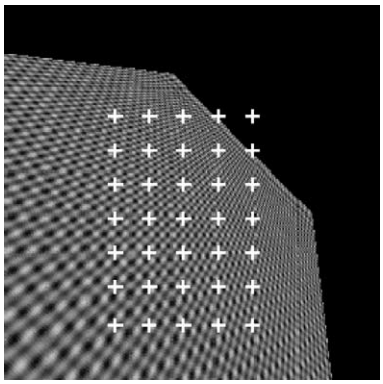


Fig. 7. Illustration of the image plane sampling procedure. Local power spectra are taken over a rectangular grid on the image plane. Each cross in the image represents the position where a local power spectrum is calculated.

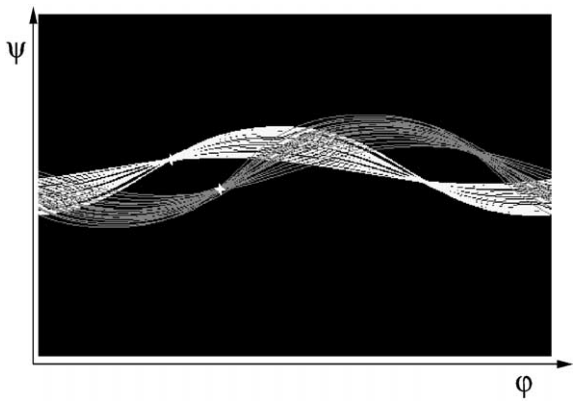



Fig. 8. Accumulating votes for vanishing points on the unit sphere.

the method is based on voting, sampling outside the texture region is not a problem. For a given plane, the presence of more than one type of texture is also not a restriction. If the texture is periodic and sufficiently regular, the spectral angles distort in the same way for every texture inside the plane region. Fig. 7 illustrates the sampling process for a rectangular region.

### 6.2. Quantisation of the voting space

Using Eq. (18), we accumulate evidence for the intersections of great circles on the unit sphere. To do this we quantise the unit sphere into accumulator cells of approximately equal area. Each great circle is traced across the unit sphere and the vote count is incremented each time it crosses a new accumulator cell. Vanishing points are located in cells which have accumulated local voting maxima. Fig. 8 illustrates this process. In the figure, the intersection points corresponding to the vanishing point locations are marked by white crosses. Once two or more intersection points are located, then the perspective pose of the plane can then be determined.

However, before we proceed we need to consider the discrete nature of the accumulation space in more detail. The unit sphere cannot be directly mapped onto a discrete rectangular accumulation space. This creates some practical difficulties for the line intersection process. For example, some lines may never intercept even when they would do so in the continuous case. Several solutions have been proposed to this problem. These include choosing different shapes for the accumulator cells and using probabilistic models of the mapping error [11–13]. There are also other ways to solve the problem. For instance, we could borrow ideas from cartography and use a Mercator projection or an equal area tessellation of the unit sphere.

We follow Barnard [15] and solve the problem by using a quadratic interpolation of the line around the neighbourhood of each point. As a practical consequence, the projected lines in the accumulation space become "thicker" or blurred around surrounding cells. Moreover, since our problem is one of shape-from-texture, we assume that there is adequate texture information available to extract a large enough sample of line directions. As a result, once the quadratic interpolation is complete, the voting process is not affected significantly by the shape of the accumulation cells.

### 6.3. Searching for the maxima

After accumulating the votes, we need to search for cells of maximum contents. The position of the cells that contain significant numbers of votes are the locations of the most likely representative vanishing points in the image. As noted in Section 2.2, we need to select at least two vanishing points to recover the parameters of the planar perspective projection. We do this by selecting the two most populated cells in the accumulation space. At this point, some practical problems with the discrete nature of the accumulation space can arise. It may occur that the same vanishing point is distributed around two or more neighbouring cells. In order to avoid this problem, and also to simplify the search, we proceed in a two-step manner. First, we find the location of the cell that contains the largest number of votes. Next, we eliminate from the accumulator the entire set of lines that intercept within a small radius of this cell. We then repeat the process again to find the location of the cell that contains the next largest number of votes.

## 7. Spectral focusing

Before we can proceed to estimate accurate perspective pose there is a chicken and egg problem which needs

to be solved. Before reliable local spectra can be estimated, there needs to be an estimate of the local distortion of the texture so that the size of the spectral window can be set. However, this local distortion is, after all, the ultimate goal of our analysis. The problem stems from the fact that if the window size is incorrectly set then the local estimate of the texture spectrum becomes de-focussed. This defocussing has serious implications for shape-from-texture. Most methods for recovering perspective pose or determining surface shape rely on finding spectral correspondences [3,18,27]. If the local spectrograms become de-focussed, then the process of matching corresponding spectral peaks can frustrated by de-localisation or merging. This problem is most marked for texture planes that are subjected to severe perspective foreshortening.

There are two important counterexamples. Garding and Lindberg [30] address the scale problem by employing a scale-space decomposition over the structural primitives. Stone and Isard [31] have a method which interleaves the adjustment of local filters for adaptive scale edge detection with the estimation of planar orientation in an iterative feedback loop. Although this provides a means of overcoming the chicken and egg nature of the estimation problem, it is couched in terms of structural textures and is sensitive to initialisation. The aim in this section, on the other hand, is to improve the accuracy of perspective pose estimation by providing a means of adaptively and locally setting the size of the spectral window. The work commences from a spectral domain analysis where we show that there is a critical window size that minimises the degree of defocussing. In order to provide a way of locally estimating this window size, we turn to the covariance matrix for the two components of the texture spectrum. Our search for the optimum window size is guided by the bias-variance structure of the covariance matrix. The size of the local spectral window is varied until the determinant of the spectral covariance matrix is minimised.

### 7.1. Unidimensional texture gradient

We now turn our attention to the effect of perspective foreshortening on the frequency contents of the texture plane. To simplify our analysis we confine our attention to the line of maximum texture gradient. This line points in the tilt direction. Without loss of generality, we can assume that the line in question is aligned with the $x$-axis in both the image and texture plane co-ordinate systems. From Eq. (2) the relationship between the two co-ordinate systems is given by

$$x_i = \frac{f x_t \cos\sigma}{h - x_t \sin\sigma}, \tag{19}$$

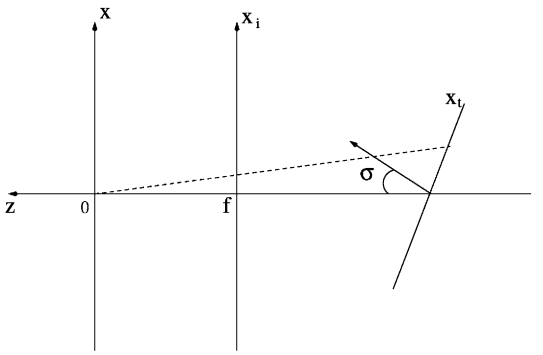



Fig. 9. Simplification of the geometric model for perspective projection.

while the inverse transformation is

$$x_t = \frac{x_i h}{x_i \sin\sigma + f\cos\sigma}. \tag{20}$$

As a result the perspective distortion of the 1-D texture signal depends only on the slant angle. The geometry of the 1-D perspective projection is shown in Fig. 9.

Now consider a simple image model in which the texture is represented by the following sinusoidal variation in intensity on the texture plane

$$I(x_t) = \cos(2\pi\omega_o x_t), \tag{21}$$

where $\omega_o$ is the frequency of the texture pattern. Using Eq. (20) the projected version of the texture pattern in the image plane is given by

$$I'(x_i) = \cos\left[2\pi\omega_o\left(\frac{x_i h}{x_i \sin\sigma + f\cos\sigma}\right)\right]. \tag{22}$$

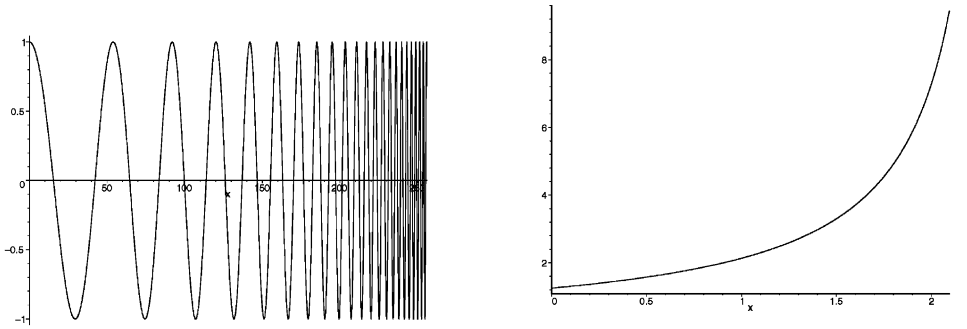
The local frequency content is a function of the parameters of the 1-D perspective projection, i.e. the height $h$, focal length $f$ and slant angle $\sigma$, together with the position on the image plane, i.e. $x_i$. This function is given by

$$\Omega(x_i) = \frac{\omega_o x_i h}{x_i \sin\sigma + f\cos\sigma}. \tag{23}$$

The texture gradient is related to the spatial derivative of the instantaneous frequency at the position $x_i$ via the formula

$$\Omega'(x_i) = \frac{\partial\Omega(x_i)}{\partial x_i} = \frac{f h\omega_o\cos\sigma}{(x_i\sin\sigma + f\cos\sigma)^2}. \tag{24}$$

This 1-D texture-gradient model is similar to a linear FM chirp waveform. The finite chirp waveform is a time-varying signal whose the frequency increases or decreases as a linear function of time. In our case, however, the perspective projection imposes a non-linear rate of




(a) (b)

Fig. 10. (a) Texture gradient as a time varying signal. (b) Time varying instantaneous frequency.

variation for the frequency with time. However, when $\tan\sigma \ll f/x_i$ then the frequency modulation follows the approximate linear form

$$\Omega'(x_i) = \frac{h\omega_0}{f\cos\sigma}\left[1 - \frac{2x_i\tan\sigma}{f} + \cdots\right]. \tag{25}$$

Fig. 10 illustrates the plot of the 1-D texture gradient represented by Eq. (22). The rate of change for the frequency in terms of time is also shown in Fig. 10.

The chirp is a non-stationary signal. Its spectral density covers a broad band of frequencies which vary in time. In other words, it has an evolutionary broad band spectrum. In order to analyse signals of this form, one approach is to minimise the observation period while maintaining a reasonable spectral resolution [32]. The Fourier spectrum of a broad band signal is continuous and covers a wide range of frequencies. At this point, it is important to point out an important limitation of the Fourier transform. When the signal is periodic and sufficiently stationary, the Fourier coefficients converge quickly. Unfortunately, this is not the case for non-stationary signals. As a result the Fourier transform itself is not satisfactory for analysing signals whose spectra vary significantly with time.

### 7.2. Windowing

When analysing signals of a non-stationary nature, it is often important to understand the correlation between the time domain and frequency domain representations of the signal. The Fourier transform itself, provides information about the frequency domain. However, the time localisation of the frequency information is essentially lost in the process of computing the Fourier transform. Another representational difficulty that can arise for non-stationary signals is related to uniqueness. Two non-stationary signals that have completely different periodicity can produce very similar spectra. As a result, Fourier analysis alone is insufficient to represent time-varying signals.

An alternative is to perform a time-spectral analysis. A non-stationary signal is divided into a sequence of time slices within which the signal is quasi-stationary. This is the idea exemplified by the short-time Fourier transform of Gabor [33]. This method uses a sliding-window Fourier transform. The relationship with the conventional Fourier transform is captured by the following definition

$$F(\omega, t_s) = \mathscr{F}\{g(t - t_s) \times f(t)\}, \tag{26}$$

where $g(t - t_s)$ is a short-time window which has a fixed width shifted along the time axis by an amount $t_s$.

This window operation allows us to locally analyse the spectral energy content of a signal over a given time interval. If the window width is sufficiently small, then its spectral content can be approximated by the instantaneous frequency at the centre point of the window. The instantaneous frequency at a specific time is given by the derivative of the angular argument of the signal at a fixed time. Eq. (22) represents a non-linear chirp waveform whose instantaneous frequency varies linearly with time. The instantaneous frequency is given by Eq. (24). Using the Fourier duality theorem, in the time domain the windowing operation is equivalent to a frequency domain convolution of the Fourier transform of the signal with its windowing function. This convolution introduces a blurring or defocusing in the Fourier domain. The amount of defocusing is inversely proportional to the width of the windowing function. It originates from the main lobe broadening introduced by the windowing function.

### 7.3. Defocusing

We will now provide an analysis of the blurring of the frequency domain spectrum which results from the windowing process. Consider the effect of windowing the signal given in Eq. (22) with a rectangular window function of width $T$. The spectral bandwidth is the range of spectral content enclosed under a specific time interval.

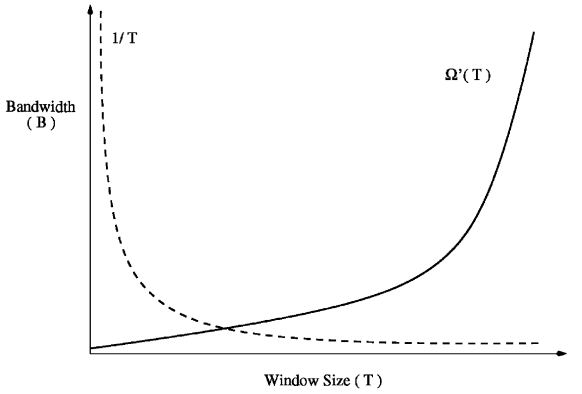



Fig. 11. Time-bandwidth plot for a rectangular window of size $T$ together with time-bandwidth plot for the 1-D texture gradient window of size $T$.

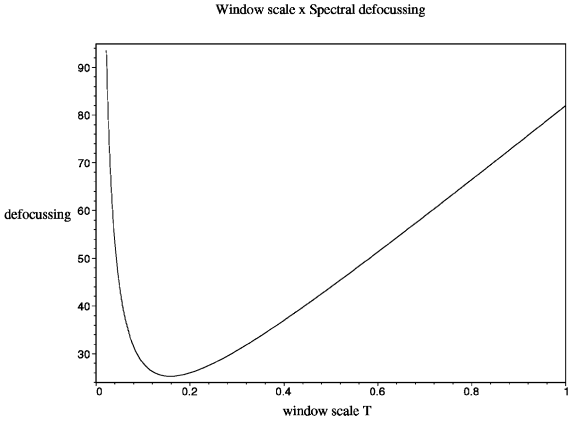



Fig. 12. Variation of the spectral defocusing in terms of the window width $T$.

Using Eq. (24), the spectral bandwidth of the 1-D texture gradient for a given window of width $T$ is given by:

$$B = 2 \times \Omega'(T) = 2\frac{fh\omega_o \cos\sigma}{(T\sin\sigma + f\cos\sigma)^2}. \tag{27}$$

In other words, the signal can be compressed into a time interval of length of $1/B$. However, we must add the excess bandwidth introduced by the windowing operation. Here there are two opposing effects. The spectral bandwidth of the signal decreases with the window size. The excess bandwidth, on the other hand, increases as the window size decreases. This is basically the tradeoff of time-localisation and frequency-resolution. These two effects are illustrated in Fig. 11. According to the figure we have the bandwidth $B$ extended by an excess due to the window main lobe contribution. The aim here is to minimise the defocusing due to the band excess while reducing the size of the window. In Fig. 12 we show a plot of the signal defocusing as a function of the window size $T$. The minimum corresponds to the optimum window size for the given bandwidth of the non-stationary signal.

This analysis suggests how we might reduce the defocusing of the Fourier spectrum by adapting the width of the local data window. The optimum size is reached when the blurring of the local spectrum is minimised. This optimum window width itself leads to an optimum estimation of the instantaneous frequency of the signal at a specific point. The idea underpinning this paper is to exploit this property to develop an unsupervised adaptive version of the short time Fourier transform. We aim to exploit the method to avoid the defocusing of the spectral representation and hence provide accurate estimation of the local spectral frequency.

### 7.4. Minimising the defocusing of the local 2-D spectral distribution

At this point we are ready to extend our unidimensional analysis to the two-dimensional analysis of textured images viewed under perspective geometry. To achieve this goal we analyse the texture gradient using local estimates of the power spectrum based on a quilt of patches on the image plane. The 2-D spectral defocusing in modelled using the covariance matrix to describe the dispersion of the spectral distribution. In this way we can determine the degree of blur in the spectral energy.

We commence by defining our local spectral estimator. In order to obtain a smooth spectral response we use the Blackman–Tukey (BT) power spectrum estimator. This is defined to be the frequency response of the windowed autocorrelation function. We employ a triangular smoothing window $w(\mathbf{X})$ due to its well documented spectral stability [34]. The spectral estimator is then

$$P(\mathbf{U}_i)_{BT} = \mathscr{F}\{r_{xx}(\mathbf{X}_i) \times w(\mathbf{X}_i)\}, \tag{28}$$

where $r_{xx}$ is the estimated autocorrelation function of the image patch. To find the optimally sized spectral window, we require a measure of spectral dispersion. Here we use the covariance matrix for the two spectral components. Formally, the matrix is defined as follows

$$\Sigma_U = E[(\mathbf{U}_i - \mu)(\mathbf{U}_i - \mu)^T] \quad \text{and} \quad \mu = E[\mathbf{U}_i]. \tag{29}$$

The eigenvalues $\lambda_1$ and $\lambda_2$ of the spectral covariance matrix are maximum and minimum values of spectral variance in an orthogonal co-ordinate system. The co-ordinate system is aligned in the direction of maximum spectral variance. If we regard the two eigenvalues as representing the two radii of a spectral ellipse, then the spectral area is equal to

$$A = \frac{\pi}{4}\lambda_1\lambda_2 = \frac{\pi}{4}|\Sigma_U|. \tag{30}$$

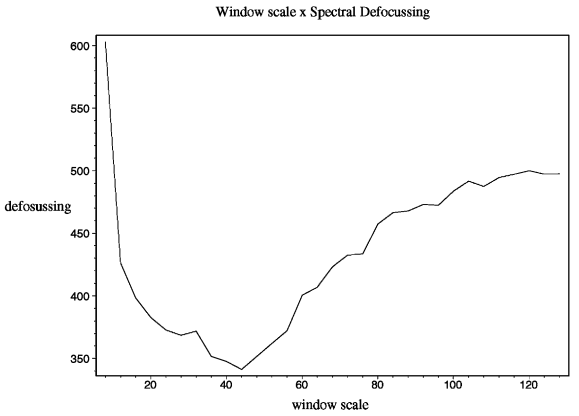



Fig. 13. Experimental plot of $|\Sigma_U|$ over increasing scale. The minimum of the distribution indicates a suitable choice for the size of the spectral analysing window.

We locate the optimal spatial domain sampling window by varying its size until the spectral area, i.e. the determinant of the spectral covariance matrix, is minimised. This in turn ensures that the dispersion, or defocusing, of the spectral moments is minimised.

### 7.5. Examples

In this section we provide some experiments with the new scale adaptation method. We commence in Fig. 13 by showing a plot of the measure of spectral dispersion, i.e. $|\Sigma_U|$ as a function of the size of the spatial window. Notice that the curve has a deep minimum. The curve also has the same gross structure as the defocusing curve shown in Fig. 12.

To proceed, we illustrate how the adaptive window responds to increasing perspectivity. In each panel of Fig. 14, the left hand image shows an artificial planar texture surface when oriented at various slant angles to the camera viewing direction. In each case there is a clear variation in the density of the texture primitives over the images. This variation occurs mainly in the tilt direction, which is also the direction of the texture gradient. Superimposed on the images are the areas of the local windows which minimise the determinant of the spectral covariance matrix. One feature to note is that each window contains approximately the same number of texture primitives. This is an important feature since the energy relationships between the image patches can only be preserved under scale consistency.

Turning our attention to the estimated spectra, Fig. 14 illustrates the defocusing produced by a poor choice of window length. The figure shows the spectra estimated at the marked positions on the planes. In each panel the top row shows the results obtained using a fixed size window, while the lower row shows the spectra obtained with the adaptive window.

We now furnish some results produced when our shape-from-texture algorithm is applied to real-world textures [35]. Fig. 15 shows three textures taken from the Brodatz album [36]. The images shown are projected at $45^\circ$ of slant. The marked patches again correspond to the optimal local spectral windows.

The final piece of experimentation aims to illustrate the utility of the new method in improving the accuracy of perspective pose estimation. Table 1 compares the estimated perspective parameters with ground truth for the artificial textures shown in Fig. 14. The main feature to note from this table is that if we use the adaptive spectral window, then we can recover accurate estimates of slant and tilt even when the planes are highly inclined. Table 2 presents the estimated perspective parameters for the three Brodatz textures shown in Fig. 15. From

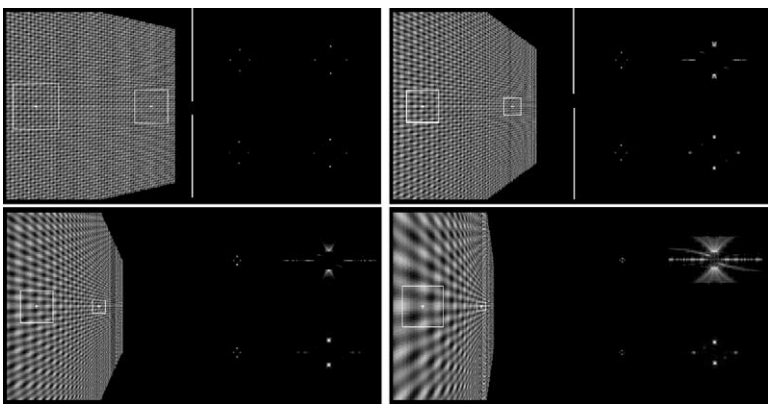
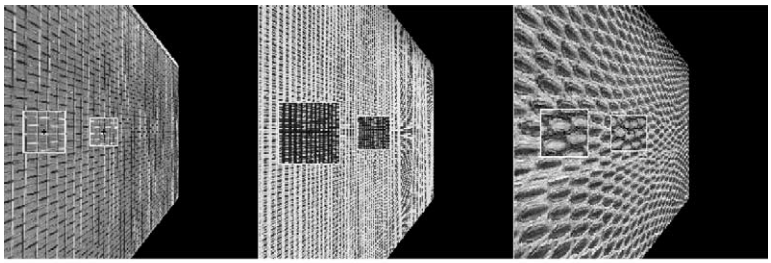


Fig. 14. Estimated scales for rectangular windows for several views of a texture plane. The corresponding spectral response is shown on the right of each figure. First row: fixed scale estimator; Second row: adaptive estimator. Increasing slant angles at $10^\circ$, $30^\circ$, $60^\circ$ and $80^\circ$.



(a) (b) (c)

Fig. 15. Estimated scales for rectangular windows for Brodatz textures at 45° slant. (a) D14—aluminium wire; (b) D21—French canvas; (c) D36—Lizard skin.

Table 1
Estimated values for slant and tilt (artificial textures)

| | Images | | Using fixed scale | | | | Using adaptive scale | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Actual values | | Estimated values | | Abs. error | | Estimated values | | Abs. error | |
| | $\sigma$ | $\tau$ | $\sigma$ | $\tau$ | $e_\sigma$ | $e_\tau$ | $\sigma$ | $\tau$ | $e_\sigma$ | $e_\tau$ |
| (a) | 10 | 0 | 12.0 | 0.0 | 2.0 | 0.0 | 9.8 | 0.0 | 0.2 | 0.0 |
| (b) | 30 | 0 | 27.5 | 0.0 | 2.5 | 0.0 | 29.5 | 0.0 | 0.5 | 0.0 |
| (c) | 60 | 0 | 62.1 | 1.0 | 2.1 | 1.0 | 59.4 | 0.0 | 0.6 | 0.0 |
| (d) | 80 | 0 | 73.4 | 2.7 | 6.6 | 2.7 | 80.4 | 0.2 | 0.4 | 0.2 |

Table 2
Estimated values for slant and tilt (Brodatz Textures)

| | Images | | Using fixed scale | | | | Using adaptive scale | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Actual values | | Estimated values | | Abs. error | | Estimated values | | Abs. error | |
| | $\sigma$ | $\tau$ | $\sigma$ | $\tau$ | $e_\sigma$ | $e_\tau$ | $\sigma$ | $\tau$ | $e_\sigma$ | $e_\tau$ |
| (a) | 45 | 0 | 41.6 | 2.0 | 3.4 | 2.0 | 46.5 | −1.0 | 1.5 | 1.0 |
| (b) | 45 | 0 | 47.0 | −0.9 | 2.0 | 0.9 | 44.5 | −1.0 | 0.5 | 1.0 |
| (c) | 45 | 0 | 73.4 | −10.9 | 28.4 | 10.9 | 55.3 | 9.0 | 10.3 | 9.0 |

these results we see improvements in the pose parameter estimates even for non-regular textures.

## 8. Experiments

In this section, we provide some results which illustrate the accuracy of planar pose estimation achievable with our shape-from-texture algorithm. This evaluation is divided into three parts. We commence by considering textures with known ground-truth slant and tilt. This part of the study is based on both synthetic textures and projected Brodatz textures [36]. The second part of our experimental study focuses on natural texture planes where the ground truth is unknown. In order to give some idea of the accuracy of the slant and tilt estimation process, we back-project the textures onto the fronto-parallel plane. Since the textures are man-made and rectilinear in nature, the inaccuracies in the estimation process manifest themselves as residual skew. Finally, we provide a sensitivity analysis to establish the operating limits of the new texture-plane pose estimation algorithm.

### 8.1. Synthetic texture planes

We commence with some examples for a synthetic regular sinusoid texture. Fig. 16 shows the synthetic texture in a number of poses. Superimposed on the textures are some of the detected lines radiating from the vanishing points. In Table 3 we list the ground truth and estimated orientation angles. Also listed is the absolute error. The

Table 3
Actual × estimated slant and tilt values (artificial images)

| | Actual | | Estimated | | Abs. Error | |
|---|---|---|---|---|---|---|
| | ($\sigma$) | ($\tau$) | ($\sigma'$) | ($\tau'$) | ($\sigma'$) | ($\tau'$) |
| (a) | 20 | 0 | 20.4 | 0.0 | 0.4 | 0.0 |
| (b) | 30 | −30 | 29.4 | −29.4 | 0.6 | 0.6 |
| (c) | 45 | 45 | 45.2 | 45.5 | 0.2 | 0.5 |
| (d) | 50 | 225 | 50.1 | 224.4 | 0.1 | 0.6 |
| (e) | 60 | 280 | 59.6 | 278.8 | 0.4 | 1.2 |
| (f) | 70 | 0 | 70.5 | −0.4 | 0.5 | 0.4 |

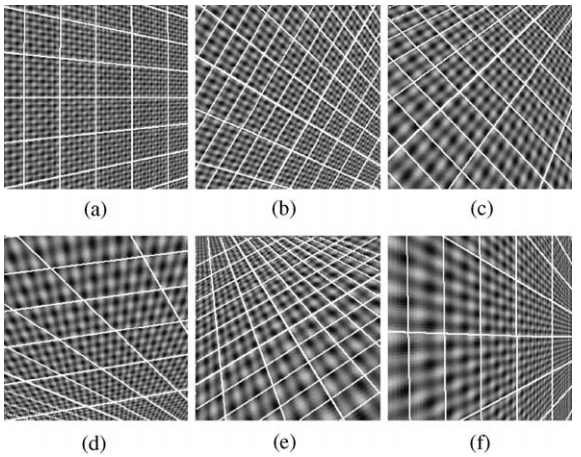



Fig. 16. Artificial textures—sinusoidal texture and detected radiating lines.

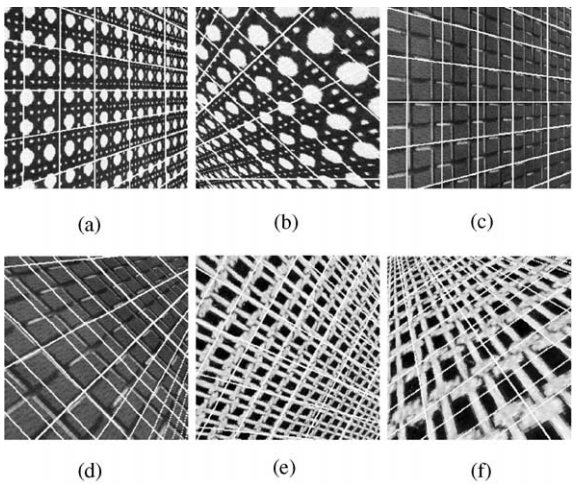



Fig. 17. Projected Brodatz textures. (a) and (b) D101; (b) and (c) D1; (d) and (e) D20.

agreement between the ground truth and estimated values is generally very good.

A second group of images is shown in Fig. 17. In this group we have taken three different texture images from the Brodatz album and have projected them onto planes of known slant and tilt. The examples here are regular natural textures of almost uniform element distribution. Superimposed on the projected textures are the estimated lines radiating from the corresponding vanishing points as given by our algorithm. Fig. 18 shows the back projection of the images onto the recovered texture plane. In most cases there is little residual perspective distortion. The values for the estimated orientation angles are listed and compared with ground truth in the Table 4. The main feature to note is that the method performs well even when the texture plane is highly inclined.

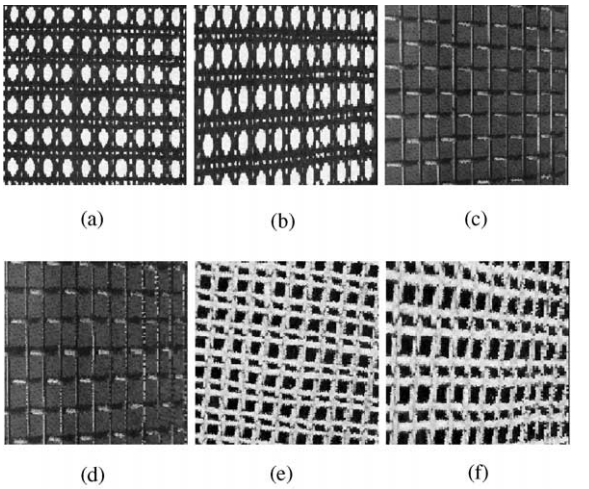



Fig. 18. Back-projected Brodatz textures. (a) and (b) D101; (b) and (c) D1; (d) and (e) D20.

### 8.2. Real world examples

This part of the experimental work focuses on real world textures with unknown ground truth. The set of real world experiments focuses on recovering the vanishing points for large man-made planar objects. The images are provided by indoor views of our lab and views of skyscrapers collected from the internet. These images contain both texture information and bounding

Table 4
Actual × estimated slant and tilt values (Brodatz textures)

| | Actual | | Estimated | | Abs. Error | |
|---|---|---|---|---|---|---|
| | $(\sigma)$ | $(\tau)$ | $(\sigma')$ | $(\tau')$ | $\sigma'$ | $\tau'$ |
| (a) | 30 | 0 | 31.0 | 0.0 | 1.0 | 0.0 |
| (b) | 50 | 225 | 51.2 | 224.1 | 1.2 | 0.9 |
| (c) | 30 | 0 | 28.1 | 0.0 | 1.9 | 0.0 |
| (d) | 45 | 45 | 42.7 | 45.3 | 2.3 | 0.3 |
| (e) | 30 | −30 | 29.6 | −32.0 | 0.4 | 2.0 |
| (f) | 60 | 120 | 58.8 | 117.8 | 1.2 | 2.2 |

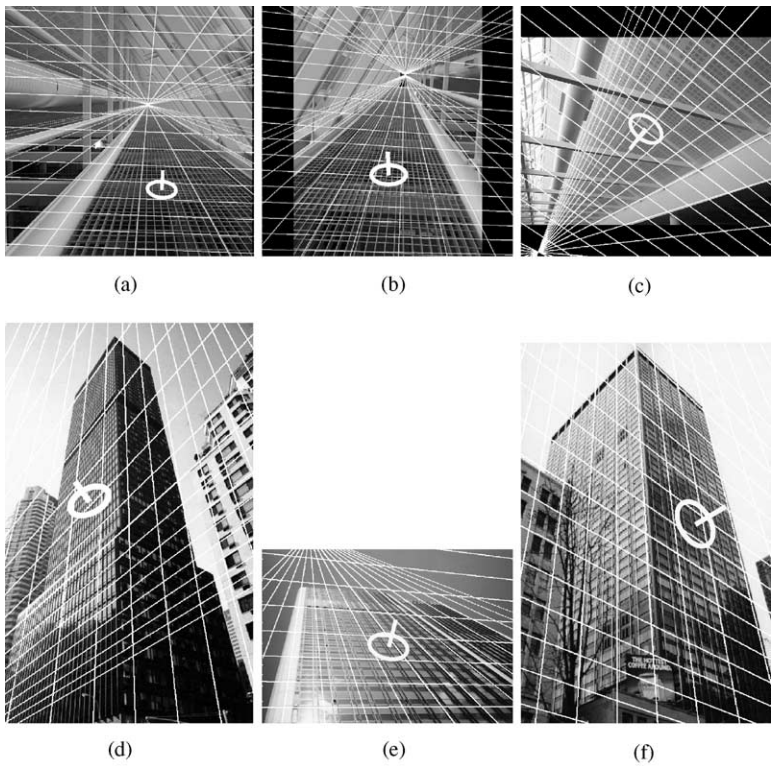



Fig. 19. Lab scenes and skyscraper images.

rectangles that can be used to confirm by visual inspection the direction of the vanishing point. These lines are not used in the analysis. For instance in the indoor scenes there are rectangular beams and gantries which converge to the vanishing point. In the skyscraper images, the sides of the planar faces can be used to confirm the accuracy of the vanishing point.

In Fig. 19 we show the reconstructed perspective planes superimposed on the original images. There are several examples of both the indoor scenes and the skyscraper images. The textures used in the analysis are quit varied. For instance, in the indoor scenes there are rectangular ventilation ducts on the roofing and webbing on the gantry. The skyscraper textures are largely composed of rectangular window patterns. The planes are visualised by drawing radial lines from the vanishing point. In all cases, the reconstructed radial lines from the vanishing point match the bounding rectangles. In other words, the reconstructions are good.

### 8.3. Effect of texture irregularity

We now turn our attention to measuring the sensitivity of the method on synthetic images. We commence by

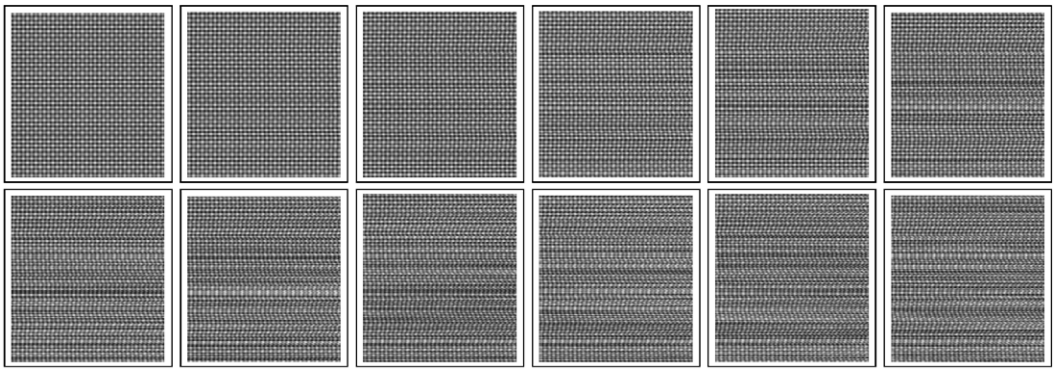


Fig. 20. Adding increased amounts of white noise to a sinusoidal texture.

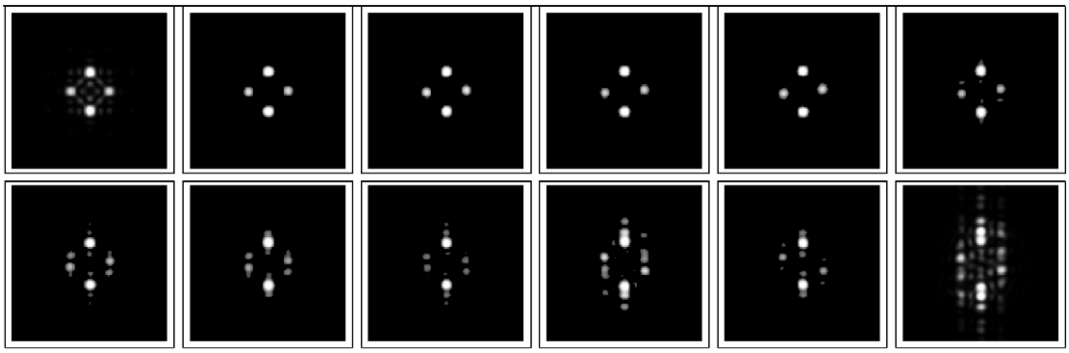


Fig. 21. Sinusoidal images: local power spectra.

assessing the ability of the method to recover reliable slant and tilt information under increasing degradation of the texture regularity. Here we have generated synthetic textured surfaces and have disturbed the texture regularity in a controlled manner.

We use the ground-truth vanishing point location to investigate the systematics of the error in the estimated plane orientation. In this way we measured the sensitivity of our method to imperfections in the regularity in the texture. In common with most of the shape-from-texture methods reported in the literature, our method is based on the assumption that local texture planes are homogeneously painted with regular textures. We have investigated with the effect of disturbing the regularity of the texture in both the frequency domain and the spatial domain.

*8.3.1. Frequency irregularity*

In the frequency domain, we have added white noise, i.e. noise with a uniform frequency distribution, of increasing amplitude to the texture. Fig. 20 shows the effect of adding increased amounts of white noise to a sinusoidal texture. Note that we add the noise to the dominant frequency component of the sinusoid. In this way we create local irregularities that will directly affect the local structure of the spectral distribution of the texture. In this part of the experiment we increase the noise level in the range of 0 to 22 per cent of the original frequency value. The degradation of the texture shows itself as ripples on the surface of the image. In order to illustrate the degradation, also in the frequency domain, in Fig. 21 we show the corresponding power spectra for each texture in Fig. 20.

In this sequence of power spectra there are two main frequency peak directions (together with their spectral reflections). The added white noise degrades the spectral information by creating extra frequency components. The noise also shifts the position of the peaks. This degradation becomes significant in Fig. 21(j) when the noise level (i.e. the ratio of the amplitude of the added white noise and the amplitude of the sinusoid) exceeds 12%.

The perspective projections of the randomised textures with slant and tilt angles of $45^{\circ}$ are shown in Fig. 22. From these images it is clear that the apparent perspectivity is sharply affected by the irregularity of the textures. For example, by visual inspection, the perspectivity becomes severely distorted in Figs. 22(g)–(m).

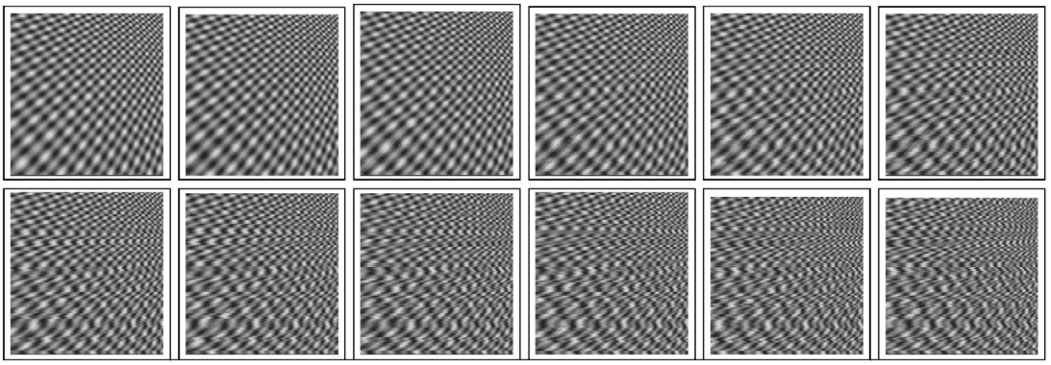


Fig. 22. Projected sinusoidal images.

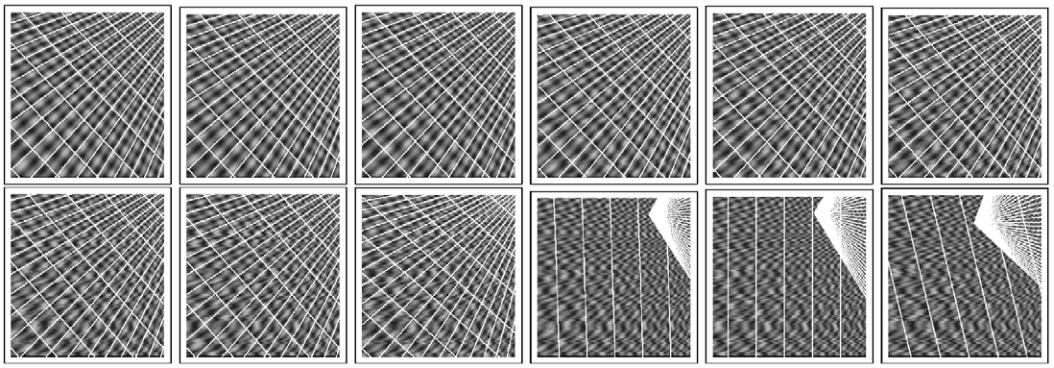


Fig. 23. Detected radiating lines.

In Fig. 23 we show the perspective poses recovered using our Hough-based vanishing point location method. In the figure we show each recovered texture plane. The set of lines radiate from the horizontal and vertical vanishing points. As expected from the observed degradation of the corresponding power spectra, our algorithm fails once the noise level reaches 12%. Figs. 23(j)–(m) illustrate the behaviour of the method after the onset of failure.

*8.3.2. Spatial location irregularity*

To complement this frequency domain study we have also investigated the effect of spatial domain texture irregularity. In the spatial domain, we have investigated the effect of randomising the positions of the texture primitives. We do this by adding a random displacement sampled from a two-dimensional circularly symmetric Gaussian distribution of zero mean and known variance. In Fig. 24, we show the effect of increasing the noise variance on the distribution of texture primitives. The corresponding power spectra for each texture in the figure are shown in Fig. 25. Again, we see the effects of increasing the irregularity in the location of the texture primitives. In the sequence, we observe that the individual peaks not only merge together, creating clouds of peaks, but also the overall frequency structure of the texture spectra degrades. This collapse of the frequency content is a result of increasing the gap spacing between the texture primitives. The overlap of the texture primitives also contributes to the decrease of the total frequency energy.

The perspective projections of these textures are shown in Fig. 26. Again, we have projected the noise corrupted, i.e. irregular, textures onto a plane whose slant and tilt angles are both equal to $45^{\circ}$. In this example, it is interesting to notice that we still perceive the sensation of perspectivity from the images in the sequence shown in Fig. 26. This may be attributable to the fact that the area gradient is invariant to the positional irregularity of the texture primitives [2,37]. The random Gaussian noise in the locations of the texture primitives does not mimic perspective distortion. As a result, the density gradient of the primitives is not affected. However, our method is based on measurements that reflect the dominant directions of the texture. These dominant directions are severely affected by the positional noise.

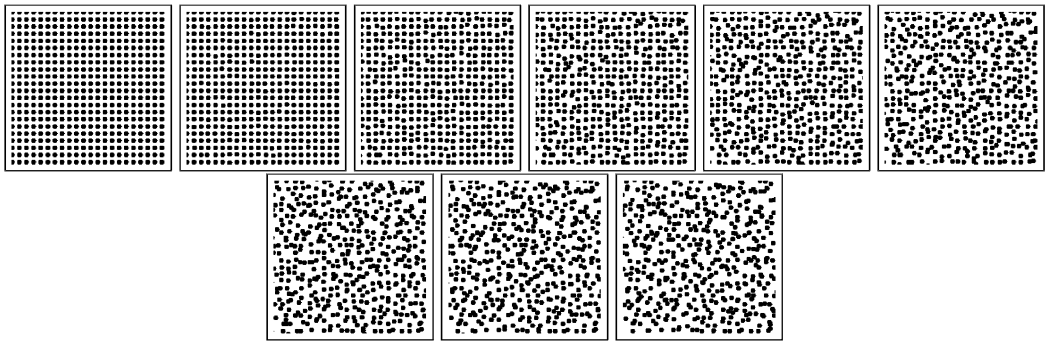


Fig. 24. Adding increased amounts of white noise to the location of circular textures.

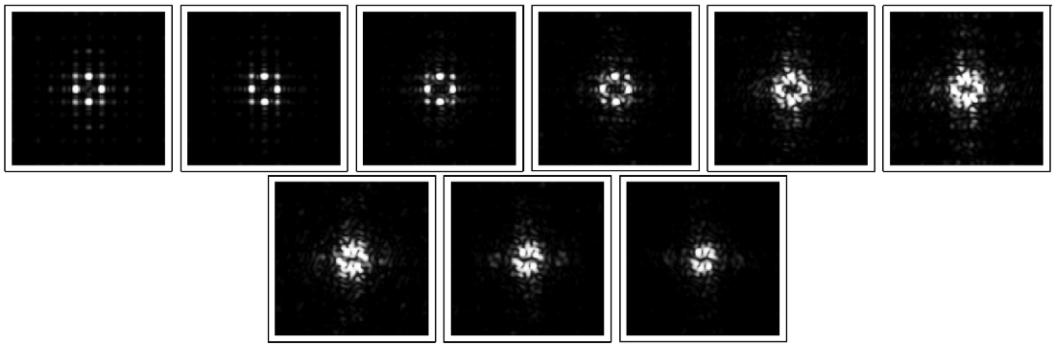


Fig. 25. Circular primitive images: local power specta.

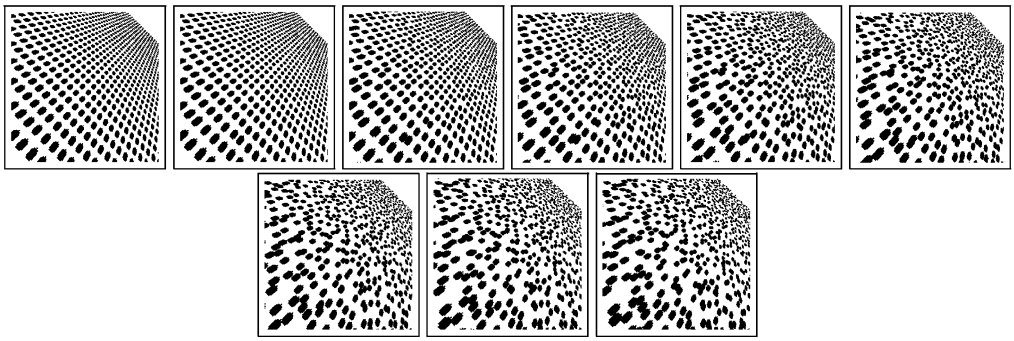


Fig. 26. Projected circular primitive images.

Fig. 27 shows the lines that radiate from the vanishing points superimposed on the texture planes. The breaking point of the algorithm occurs once the ratio of the standard deviation of the displacement error to the inter-point distance becomes larger than 10%.

In Fig. 28 we plot the effect of the spatial domain and frequency domain noise on the error in the recovered vanishing point locations. The plot shows the measured error in the horizontal vanishing point location for the sinusoidal textures shown in Fig. 22. Here the method degrades once the amplitude of the white noise exceeds 12% of the amplitude of the original sinusoidal frequency. Fig. 28(b) repeats this experiment for the spatial-domain texture shown in Fig. 26. The plot shows the vanishing point location error as a function of the standard deviation in the random spatial displacement error. Here the method recovers good estimates of the vanishing point location provided that the ratio of the standard deviation

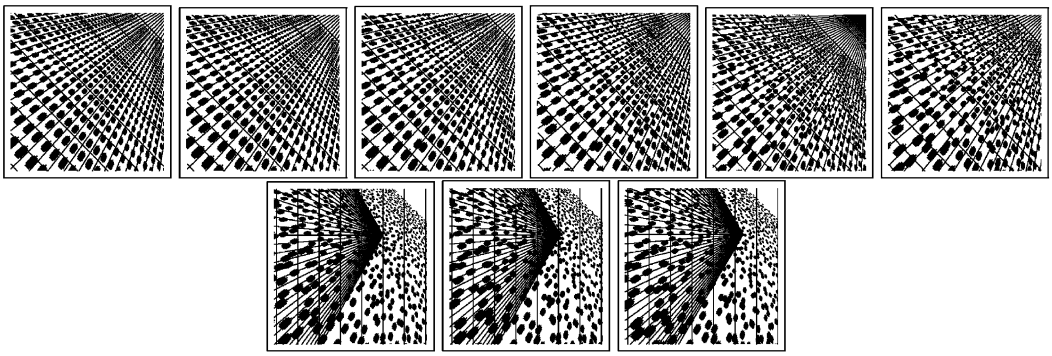


Fig. 27. Estimated radiating lines.

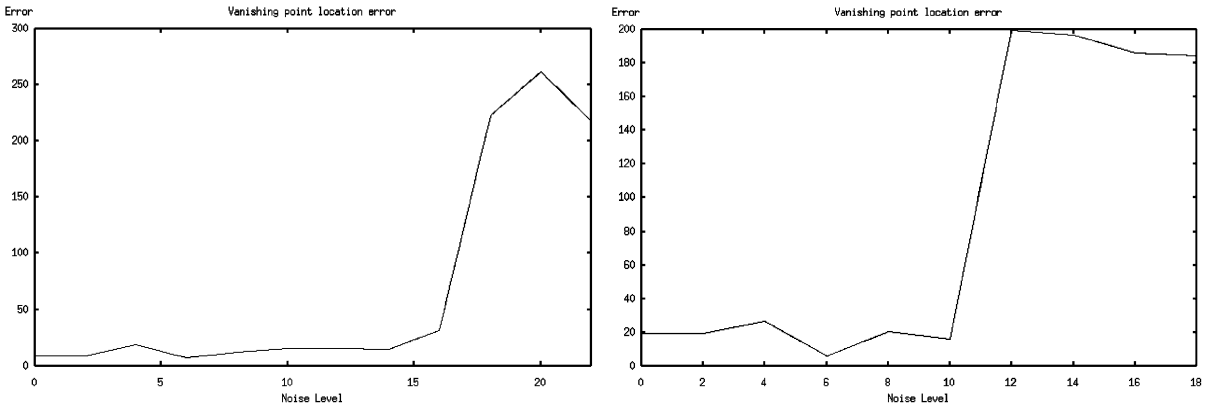



Fig. 28. Sensitivity plots.

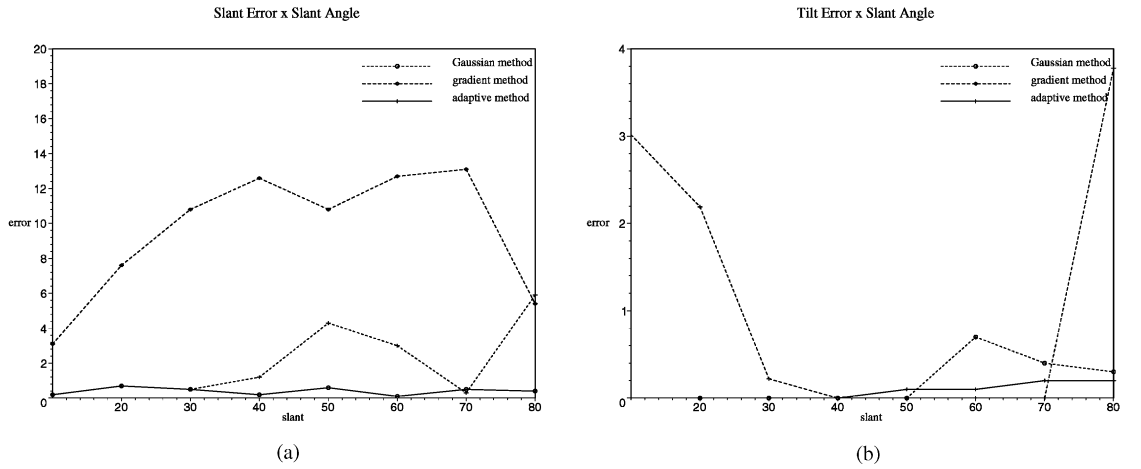
of the displacement error to the inter-point distance does not exceed 10%.

### 8.4. Sensitivity analysis

Finally, we provide some comparison between our new method and two alternatives. The first of these is a conventional gradient-based method [35]. The second method uses a fixed-size window together with the unit sphere representation. In this analysis we investigate the behaviour of the estimated orientation with varying slant and tilt angles. We use a sinusoidal image texture of the sort used in Fig. 16.

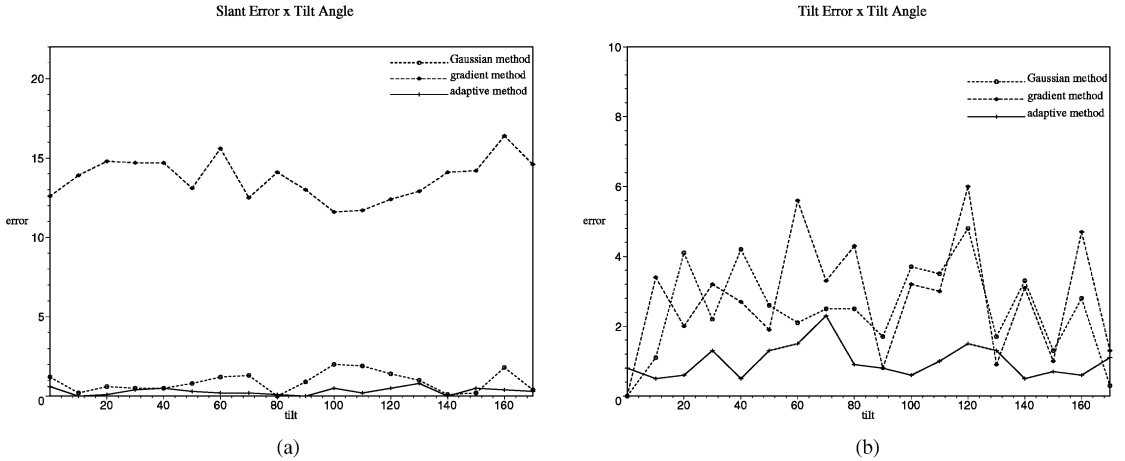
We commence by studying the errors for the slant and tilt angle estimation when only the slant angle varies. The graphs in Fig. 29 plot the slant and tilt errors when the orientation varies from $10^\circ$ to $80^\circ$ and the tilt angle remains constant at $0^\circ$. The solid line is the result of applying our new method for perspective pose estimation (this algorithm uses both the unit sphere and an adaptive spectral window). The dotted curve is the result obtained when the size of the spectral window is kept constant. Finally, the dashed curve is the result of using an image-plane texture gradient method. Fig. 29(a), shows the slant error as a function of the slant angle. The main feature to note from this plot is as follows. Irrespective of whether the adaptive or fixed windows are used, at small slant angles the search of the unit-sphere outperforms gradient-based search of the texture plane. Specifically, the unit sphere gives slant errors of a few tenths of a degree, while texture gradient gives errors of the order of a few degrees. When the adaptive window is used, then the slant error is approximately constant. However, at large slant angles the fixed-size window gives a slant error of about $3^\circ$ while the texture-plane method has an average slant error of $9.5^\circ$.

Fig. 29(b) shows the tilt error as a function of slant angle. In the case of search of the unit sphere with an adaptive spectral window, the error increases slowly with slant angle. However, it is never more than a few tenths of a degree. When a fixed size window is used, then the same pattern emerges. However, the error is now slightly larger. Finally, the texture-plane method gives very poor errors for both small and large slant angles. The poor performance at small slant angles is attributable to problems associated with vanishing texture gradient.




(a) (b)

Fig. 29. Slant/tilt error plot for both methods. Slant varying from $10^{\circ}$ to $80^{\circ}$. (a) Slant error; (b) Tilt error.




(a) (b)

Fig. 30. Slant/tilt error plot for both methods. Tilt varying from $0^{\circ}$ to $170^{\circ}$. (a) Slant error; (b) Tilt error.

In Fig. 30 we show the estimation error under varying tilt and fixed slant angle. The tilt varies over the range from $0^{\circ}$ to $170^{\circ}$, whilst the slant angle remains constant at $40^{\circ}$. In Fig. 30(a) we show the slant error. When the unit sphere representation is used, the error is typically a few tenths of a degree. Since we are working at moderate slant angle, the error obtained with adaptive window is consistently slightly lower than that obtained with a fixed size window. By contrast, the average error for the gradient method is $13^{\circ}$. In Fig. 30(b) we plot the tilt error as a function of tilt angle. There is little structure in this plot. Moreover, there is little to distinguish the three methods. The average tilt error for the texture plane method is $2.8^{\circ}$. The use of the unit sphere with a fixed size spectral window results in an average tilt error of $2.4^{\circ}$. When the adaptive window is used, the average tilt error is $1^{\circ}$.

## 9. Conclusions

We have described an algorithm for estimating the perspective pose of textured planes by projecting spectral information onto the unit sphere. We exploit the fact that the local spectral components are oriented in the direction of vanishing points. As a result each estimated spectral component can be mapped onto a great circle of the unit sphere. Vanishing points are characterised by locations at which several great circles intercept. Based on this observation, we pose the problem of estimating perspective pose as that of searching for accumulator cells of maximum contents on the unit sphere. The method is illustrated to operate effectively on both synthetic imagery with known ground truth and on a wide variety of real-world textured planes. One advantage of the method is that it does not rely on potentially unreliable estimates

of texture gradient to constrain the tilt angle. Moreover, we overcome problems of poor parameter estimation at large slant angles by adapting the size of the spectral window using the spectral covariance matrix. As a result the estimated slant and tilt angles are more accurately determined.

There are a number of ways in which the ideas presented in this paper can be extended. Our next step is to consider how the method can be applied to curved surfaces, using the method to estimate local slant and tilt parameters. Studies aimed at developing these ideas are in hand and will be reported in due course.

## References


[1] D. Marr, Vision: A computational investigation into the human representation and processing of visual information, Freeman, New York, 1982.

[2] J.J. Gibson, The Perception of the Visual World, Houghton Miffin, Boston, MA, 1950.

[3] J. Krumm, S.A. Shafer, Texture segmentation and shape in the same image, IEEE International Conference on Computer Vision, Boston, 1995, pp. 121–127.

[4] B.J. Super, A.C. Bovik, Shape from texture using local spectral moments, IEEE Trans. Pattern Anal. Mach. Intell. 17 (4) (1995) 333–343.

[5] J.R. Kender, Shape from texture: an aggregation transform that maps a class of texture into surface orientation, Sixth IJCAI, Tokyo, 1979, pp. 475–480.

[6] J.S. Kwon, H.K. Hong, J.S. Choi, Obtaining a 3D orientation of projective textures using a morphological method, Pattern Recognition 29 (1996) 725–732.

[7] J. Garding, Shape from texture for smooth curved surfaces, European Conference on Computer Vision, 1992, pp. 630–638.

[8] E. Ribeiro, E.R. Hancock, Improved orientation estimation for texture planes using multiple vanishing points, Pattern Recognition 10 (2000) 1599–1610.

[9] E. Ribeiro, E.R. Hancock, Estimating the 3D orientation from texture planes using local spectral analysis, Image and Vision Comput. 18 (8) (2000) 619–631.

[10] J.A. Shufelt, Performance evaluation and analysis of vanishing point detection techniques, IEEE Trans. Pattern Anal. Mach. Intell. 21 (3) (1999) 282–288.

[11] P.L. Palmer, A.T. Tai, An optimised vanishing point detector, British Machine Vision Conference, Guildford, 1993, pp. 529–538.

[12] B. Brillaut-O'Mahony, New method for vanishing point detection, CVGIP: Image Understand. 54 (2) (1991) 289–300.

[13] E. Lutton, H. Maitre, J. Lopez-Krahe, Contribution to the detection of vanishing points using hough transform, IEEE Trans. Pattern Anal. Mach. Intell. 16 (4) (1996) 430–438.

[14] G.F. Mclean, D. Kotturi, Vanishing point detection by line clustering, IEEE Trans. Pattern Anal. Mach. Intell. 17 (11) (1995) 1090–1095.

[15] S.T. Barnard, Interpreting perspective images, Artif. Intell. 21 (1983) 435–462.

[16] Y. Ohta, K. Maenobu, T. Sakay, Obtaining surface orientation from texels under perspective projection, Seventh IJCAI, Vancouver, 1981, pp. 746–751.

[17] J. Krumm, S.A. Shafer, Shape from periodic texture using spectrogram, IEEE Conference on Computer Vision and Pattern Recognition, Urbana, IL, 1992, pp. 284–289.

[18] B.J. Super, A.C. Bovik, Planar surface orientation from texture spatial frequencies, Pattern Recognition 28 (5) (1995) 729–743.

[19] J. Aloimonos, Perspective approximations, Image Vision Comput. 8 (3) (1990) 179–192.

[20] J. Krumm, S.A. Shafer, Local spatial frequency analysis of image texture, IEEE International Conference on Computer Vision, Tokyo, 1990, pp. 354–358.

[21] C.J. Poelman, T. Kanade, A paraperspective factorization method for shape and motion recovery, PAMI 19 (3) (1997) 206–218.

[22] T.M. Apostol, Calculus, 2nd Edition, Wiley Series, New York, 1969.

[23] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, Addison-Wesley, Reading, MA, 1993.

[24] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, USA, 1973.

[25] J. Malik, R. Rosenholtz, Recovering surface curvature and orientation from texture distortion: a least squares algorithm and sensitive analysis, Lectures Notes in Computer Science—ECCV'94, Vol. 800 1994, 353–364.

[26] R.N. Bracewell, K.-Y. Chang, A.K. Jha, Y.-H. Wang, Affine theorem for two-dimensional fourier transform, Electron. Lett. 29 (3) (1993) 304.

[27] J. Malik, R. Rosenholtz, A differential method for computing local shape-from-texture for planar and curved surfaces, IEEE Conference on Vision and Pattern Recognition, New York, 1993, pp. 267–273.

[28] K. Ikeuchi, Shape from regular patterns, Artif. Intell. 22 (1984) 49–75.

[29] A.C. Bovik, M. Clark, W.S. Geisler, Multichannel texture analysis using localized spatial filters, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1) (1990) 55–73.

[30] J. Garding, T. Lindeberg, Direct computation of shape cues using scale-adapted spatial derivatives operators, Int. J. Comput. Vision 17 (2) (1994) 163–191.

[31] J.V. Stone, S.D. Isard, Adaptive scale filtering: A general method for obtaining shape from texture, IEEE Trans. Pattern Anal. Mach. Intell. 17 (7) (1995) 713–718.

[32] W.G. Carrara, R.S. Goodman, R.M. Majewski, Spotlight Synthetic Aperture Radar: Signal Processing Algorithms, Artech House, USA, 1995.

[33] D. Gabor, Theory of communication, J. Inst. Electron. Eng. 93 (1946) 429–457.

[34] S.M. Kay, Modern Spectral Estimation: Theory and Application, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[35] E. Ribeiro, E.R. Hancock, 3-D planar orientation from texture: estimating vanishing point from local spectral analysis, IX British machine Vision Conference, September 1998, pp. 326–335.

[36] Phil Brodatz, Textures: A Photographic Album for Artists and Designers, Dover, New York, 1966.

[37] K.A. Stevens, The information content of texture gradients, Biol. Cybernet. 42 (1981) 95–105.

**About the Author**—ERALDO RIBEIRO recently completed a D.Phil. degree in Computer Vision in the Department of Computer Science at the University of York. Prior to this, he gained his Master of Science Degree with distinction in Computer Science (Image Processing) at the Federal University of Sao Carlos (UFSCar-SP), Brazil in 1995. His first degree is in Mathematics from the Catholic University of Salvador, Brazil, 1992. His research interests are in shape from texture techniques and 3-D scene analysis.

**About the Author**—EDWIN HANCOCK studied Physics as an undergraduate at the University of Durham and graduated with honours in 1977. He remained at Durham to complete a Ph.D. in the area of high energy physics in 1981. Following this he worked for 10 years as a researcher in the fields of high-energy nuclear physics and pattern recognition at the Rutherford–Appleton Laboratory (now the Central Research Laboratory of the Research Councils). During this period, he also held adjunct teaching posts at the University of Surrey and the Open University. In 1991, he moved to the University of York as a lecturer in the Department of Computer Science. He was promoted to Senior Lecturer in 1997 and to Reader in 1998. In 1998 he was appointed to a Chair in Computer Vision.

Professor Hancock now leads a group of some 15 faculty, research staff and Ph.D. students working in the areas of computer vision and pattern recognition. His main research interests are in the use of optimisation and probabilistic methods for high and intermediate level vision. He is also interested in the methodology of structural and statistical pattern recognition. He is currently working on graph-matching, shape-from-X, image data-bases and statistical learning theory. His work has found applications in areas such as radar terrain analysis, seismic section analysis, remote sensing and medical imaging. Professor Hancock has published some 60 journal papers and 200 refereed conference publications. He was awarded the Pattern Recognition Society medal in 1991 for the best paper to be published in the journal Pattern Recognition. The journal also awarded him an outstanding paper award in 1997.

Professor Hancock is a member of the Editorial Boards of the journals IEEE Transactions on Pattern Analysis and Machine Intelligence, and, Pattern Recognition. He has also been a guest editor for special editions of the journals Image and Vision Computing and Pattern Recognition, and he is currently a guest editor of a special edition of IEEE Transactions on Pattern Analysis and Machine Intelligence devoted to energy minimisation methods in computer vision. He has been on the programme committees for numerous national and international meetings. In 1997, he established a new series of international meetings on energy minimisation methods in computer vision and pattern recognition.

In 2000, he became a Fellow of the International Association for Pattern Recognition.