PERGAMON

Pattern Recognition 33 (2000) 1599–1610

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Improved orientation estimation for texture planes using multiple vanishing points

Eraldo Ribeiro[1], Edwin R. Hancock*

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

Received 11 February 1999

Abstract

Vanishing point locations play a critical role in determining the perspective orientation of textured planes. However, if only a single vanishing point is available then the problem is undetermined. The tilt direction must be computed using supplementary information such as the texture gradient. In this paper we describe a method for multiple vanishing point and, hence complete perspective pose estimation, which obviates the need to compute the texture gradient. The method is based on local spectral analysis. It exploits the fact that spectral orientation is uniform along lines that radiate from vanishing points on the image plane. We experiment with the new method on both synthetic and real-world imagery. This demonstrates that the method provides accurate pose angle estimates, even when the slant angle is large. © 2000 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Perspective pose recovery; Planar shape from texture; Spectral analysis; Multiple vanishing points

1. Introduction

The perspective foreshortening of surface patterns is an important one for the recovery of surface orientation from 2D images [1,2]. Broadly speaking there are two routes to recover the parameters of perspective projection for texture patterns. The first of these is to estimate the texture gradient [3,4]. Geometrically, the texture gradient determines the tilt direction of the plane in the line-of-sight of the observer and its magnitude determines the slant angle of the plane. A more direct and geometrically intuitive alternative route to the local slant and tilt parameters of the surface is to estimate the whereabouts of vanishing points [5,6]. When only a single vanishing point is available, the direction of the texture gradient is still a pre-requisite since the surface orientation parameters can only be determined provided that the tilt direction is known. However, if two or more vanishing points are available, then not only can the slant and tilt be determined uniquely, they can also be determined more accurately.

Unfortunately, the location of a single vanishing point from texture distribution is not itself a straightforward task. If direct analysis is being attempted in the spatial domain, then the tractability of the problem hinges on the regularity and structure of the texture primitives [5,6]. Moreover, multiple vanishing point detection may be even more elusive. It is for this reason that frequency domain analysis offers an attractive alternative. The main reason for this is that the analysis of spectral moments can provide a convenient means of identifying the individual radial patterns associated with multiple vanishing points.

1.1. Related literature

To set the work reported in this paper in context we provide an overview of the related literature. We commence by considering texture gradient methods for shape from texture. Broadly speaking, there are two ways in which texture gradients can be used for shape estimation.

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

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

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

PII: S0031-3203(99)00148-X

The first of these is to perform a structural analysis of pre-segmented texture primitives using the geometry of edges, lines or arcs [7–9]. The attractive feature of such methods is the way in which their natural notion of perspective appeals to the psychophysics of visual perception. Unfortunately, structural approaches are heavily dependant on the existence of reliable methods for extracting the required geometric primitives from the raw texture images. Moreover, they cannot cope with scenes in which there are no well-defined texture elements.

As mentioned above, the second approach is to cast the problem of shape-from-texture in the frequency domain [10–13]. The main advantage of the frequency domain approach is that it does not require an image segmentation as a pre-requisite. For this reason it is potentially more robust than its structural counterpart. Bajcsy and Lieberman [14] were among the first to use frequency analysis in order to recover shape from texture. Their work is based on the analysis of the frequency energy gradient for oblique views of outdoor images. Subsequently, the frequency domain has been exploited to provide both local and global descriptions of the perspective distortion of textured surfaces. Specifically, Super and Bovik [11] describe texture information in terms of local spectral frequencies. Perspective pose is recovered by back-projecting the spectral energy to minimise a variance criterion. Stone [15] has proposed an iterative method based on a similar local frequency decomposition. The novel feature is to include feedback into the back-projection process, whereby the local frequency filters undergo affine transformation. In this way the method is able to improve the spectral analysis for planes at high slant angles by using an adaptive scale. Underpinning these latter two methods is the local affine analysis of the spectral decomposition of textures under global perspective projection. In this respect, there are similarities with the work of Malik and Rosenholtz [16] which provides an affine quilting for neighbouring texture patches.

Alternatively, planar orientation can be recovered provided that vanishing points in the image plane can be detected. Since most textures are normally rich in geometric distortion when viewed under perspective geometry, vanishing points can be detected based on the projected texture and used to solve for planar surface orientation [5,6]. For example, Kender has proposed an aggregation transform which accumulates edges directions of the texture primitives in order to estimate a vanishing point for planar surfaces [5]. Kwon et al., on the other hand have employed mathematical morphology in order to estimate planar surface orientation from the location of a single vanishing point [6]. These methods, however, approach the texture analysis in a structural manner. They are hence limited by the inherent drawbacks mentioned above.

1.2. Paper outline

The aim of the work reported in this paper is to combine the advantages of the local spectral analysis of textures with the geometry of vanishing point detection in order to recover planar surface orientation. To achieve this goal, we model the texture content in terms of the local spectral frequency. This provides a more general representation of texture. It is therefore both more flexible and more stable to local texture variations. However, rather then using the radial frequency [11,14], we use the angular information provided by the local spectral distribution.

Our aims are twofold. Firstly, we introduce the prerequisites for our study by providing a detailed review of the properties of the local spectral moments under perspective geometry. We commence by considering the local distortions of the spectral moments. Here, we follow Super and Bovik [11] and Krumm and Shafer [10] by using the Taylor expansion to make a local linear approximation to the transformation between texture plane and image plane. In other words, we identify the affine transformation that locally approximates the global perspective geometry. With this local approximation to hand, we follow Malik and Rosenholtz [13] by applying Bracewell's [17] affine Fourier theorem to compute the local frequency domain distortions of the spectral distribution. Based on this analysis we make a novel contribution and show that lines of uniform spectral orientation radiate from the vanishing point. The practical contribution in this paper is to exploit this property to locate multiple vanishing points. We provide an analysis to show how the slant and tilt parameters can be recovered from a pair of such points. We illustrate the new method for estimating plane orientation on a variety of real-world and synthetic imagery. This study reveals that the method is accurate even when the slant angle becomes large.

The outline of this paper is as follows. In Section 2 we review the perspective geometry of texture planes. Section 3 shows how the vanishing point position is related to the slant and tilt angles of the texture plane. Details of the projective distortion of the power spectrum are presented in Section 4. These three sections provide the mathematical prerequisites for our study. They provide a synopsis of existing results widely used for shape-from-texture. For instance, Super and Bovik [11] and Krumm and Shafer [10] have both made use of locally affine approximations of perspective geometry. The relationship between perspective pose parameters and vanishing point position is nicely described in the textbook of Haralick and Shapiro [18]. The Fourier analysis of local spectra under affine projection is exploited by Malik and Rosenholtz [13] in their work on curved surface analysis. Section 5 develops the novel idea underpinning the paper, namely, that lines radiating from vanishing points

have a uniform spectral angle. Section 6 experiments with the new method for estimating perspective pose on both synthetic and real-world images, Finally, Section 7 offers some conclusions.

2. Geometric modelling

We commence by reviewing the projective geometry for the perspective transformation of points on a plane. 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 as shown in Fig. 1. Suppose that the texture plane is a distance h from the camera which has focal length $f < 0$. Consider two corresponding points that have co-ordinates $\mathbf{X}_t = (x_t, y_t)^{\mathrm{T}}$ on the texture plane and $\mathbf{X}_i = (x_i, y_i)^{\mathrm{T}}$ on the image plane. The perspective transformation between the two co-ordinate systems is

$$\mathbf{X}_i = T_p \mathbf{X}_t, \tag{1}$$

where T_p is the perspective transformation matrix. We represent the orientation of the viewed surface plane using the slant angle σ and tilt angle τ. This parametrisation is a natural way to model local surface orientation. Fig. 2 illustrates the slant and tilt model. For a given plane, considering a viewer-centred representation, the slant is the angle between viewer sight line and the normal vector of the plane. The tilt is the angle of rotation of the normal vector around the sight line axis. The elements of the transformation matrix T_p can be computed using the slant angle σ and tilt angle τ in the following manner:

$$T_p = \frac{f \cos\sigma}{h - x_t \sin\sigma} \begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\cos\sigma} \end{bmatrix}. \tag{2}$$

The perspective transformation in Eq. (2) represents a non-linear geometric distortion of a surface texture pattern onto an image plane pattern. Unfortunately, the non-linear nature of the transformation makes Fourier domain analysis of the texture frequency distribution somewhat intractable. In order to proceed, we therefore derive a local linear approximation to the perspective transformation. However, it should be stressed that the global quilting of the local approximations preserves the perspective effects required for the recovery of shape-from-texture. With this linear model, the perspective distortion can be represented as shown in Fig. 3. In the

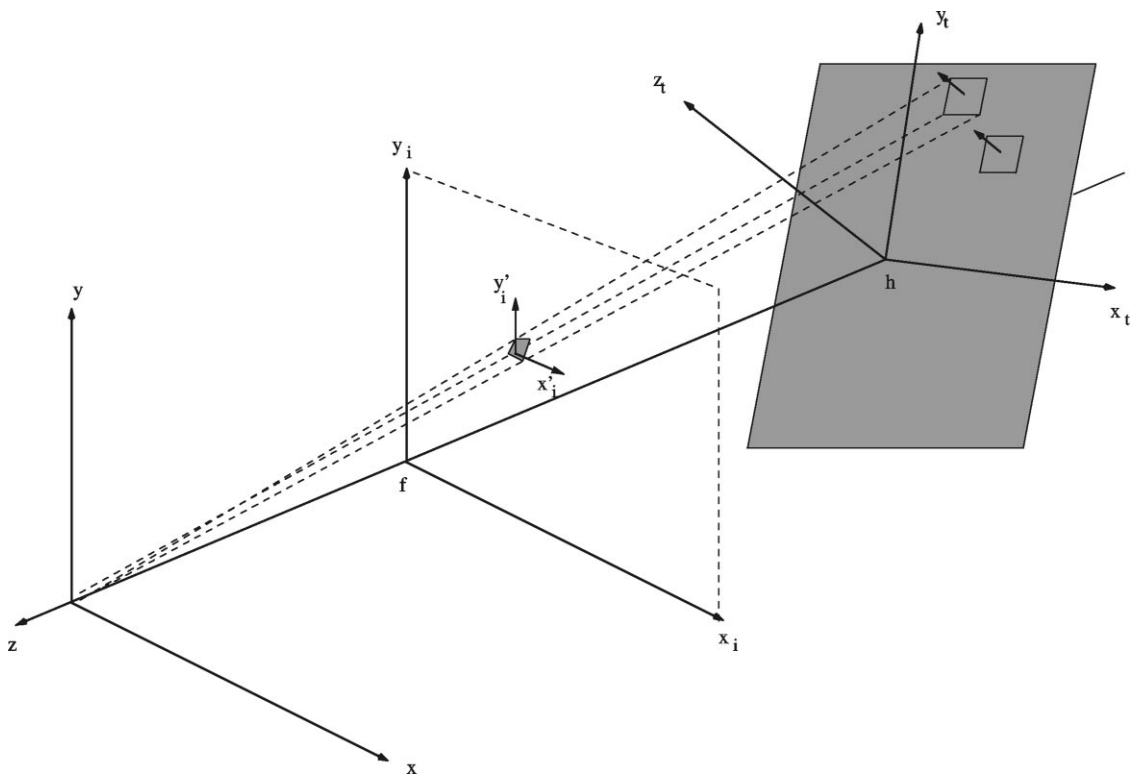

Fig. 1. Perspective projection of a planar surface onto an image plane. The projection of a local patch over the texture plane onto the image plane is also shown.

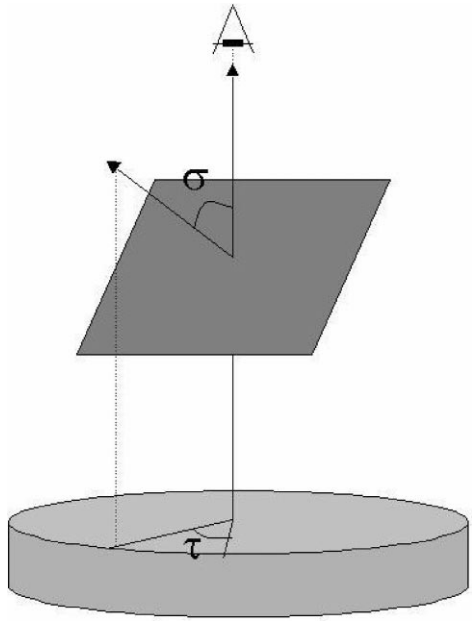

Fig. 2. Slant and tilt parametrisation of the plane orientation.

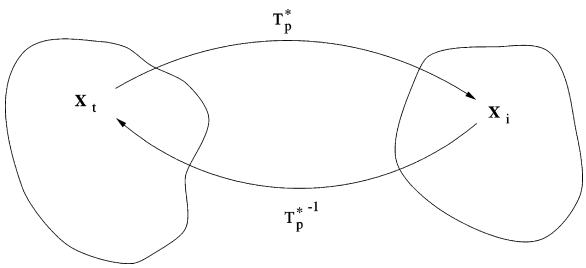

Fig. 3. Perspective mapping of X_t to X_i where T_p^* is a linear version of the perspective projection. T_p^* is an affine transformation between X_t and X_i.

diagram, T_p^* is the linear version of the perspective transformation given by Eq. (2).

Linear approximations of the local spectral distortion due to perspective projection have been employed by several authors in spectral shape-from-texture. Super and Bovik [11] derive their geometric model in terms of the instantaneous frequency at a point of the image plane. Krumm and Shafer [10] use a first-order Taylor series approximation of the perspective projection at a local point. Malik and Rosenholtz [16] employ a differential approximation of the local spectral perspective distortion.

We follow Krumm and Shafer [10] and linearize T_p using a first-order Taylor formula. Our transformation, however, is given in terms of the slant and tilt angles instead of the normal vector components p and q. Let (xo_t, yo_t, h) be the origin or expansion point of the local co-ordinate system of the resulting affine transformation. This origin projects to the point (xo_i, yo_i, f) on the image plane. We denote the corresponding local co-ordinate system on the image plane by $\mathbf{X}_i' = (x_i', y_i', f)$ where $x_i = x_i' + xo_i$ and $y_i = y_i' + yo_i$. The linearised version of T_p in Eq. (2) is obtained through the Jacobian $J(.)$ of $\mathbf{X}_i$ where each partial derivative is calculated at the point $\mathbf{X}_i' = \mathbf{0}$. After the necessary algebra, the resulting linear approximation is

$$T_p^* = J(\mathbf{X}_i)|_{(\mathbf{X}_i'=0)} = J(T_p\ \mathbf{X}_t)|_{(\mathbf{X}_i'=0)}, \tag{3}$$

where

$$J(\mathbf{X}_i) = \begin{bmatrix} \dfrac{\partial}{\partial x_i'}\, x_t(x_i', y_i') & \dfrac{\partial}{\partial y_i'}\, x_t(x_i', y_i') \\ \dfrac{\partial}{\partial x_i'}\, y_t(x_i', y_i') & \dfrac{\partial}{\partial y_i'}\, y_t(x_i', y_i') \end{bmatrix}. \tag{4}$$

Rewriting T_p^* is terms of the slant and tilt angles we have

$$T_p^* = \frac{\Omega}{hf\cos\sigma} \begin{bmatrix} xo_i \sin\sigma + f\cos\tau\cos\sigma & -f\sin\tau \\ yo_i \sin\sigma + f\sin\tau\cos\sigma & f\cos\tau \end{bmatrix}, \tag{5}$$

where $\Omega = f\cos\sigma + \sin\sigma(xo_i\cos\tau + yo_i\sin\tau)$. Hence, T_p^* depends on the expansion point (xo_i, yo_i) which is a constant.

The transformation T_p^* in Eq. (5) operates from the texture plane to the image plane. Later on when we come to consider the Fourier transform from the observed texture distribution on the image plane to that on the texture plane, it is the inverse of the transpose transformation matrix, i.e. $(T_p^{*-1})^{\mathrm{T}}$ which will be of interest. The matrix is given by

$$(T_p^{*-1})^{\mathrm{T}} = \frac{h\cos\sigma}{\Omega^2} \begin{bmatrix} f\cos\tau & -yo_i\sin\sigma - f\sin\tau\cos\sigma \\ f\sin\tau & xo_i\sin\sigma + f\cos\tau\cos\sigma \end{bmatrix}. \tag{6}$$

3. Vanishing points

The net effect of the global perspective transformation is to distort the viewer-centred texture pattern in the direction of the vanishing point $\mathbf{V} = (x_v, y_v)^{\mathrm{T}}$ in the image plane. Suppose that the object-centred texture pattern consists of a family of parallel lines which are oriented in the direction of the vanishing point. When transformed into the image-centred co-ordinate system, this family of lines can be represented using the following set of parametric equations [18]:

$$L = \{\mathbf{X}_s | \mathbf{X}_s = \mathbf{A} + \lambda\mathbf{B}\}, \tag{7}$$

where λ is the parameter of the family. The three constants forming the vector $\mathbf{B} = (b_1, b_2, b_3)^{\mathrm{T}}$ are the direction cosines for the entire family L. The individual lines in

L are each parametrised by the vector $\mathbf{A} = (a_1, a_2, a_3)^{\mathrm{T}}$. The co-ordinates of the vanishing point $\mathbf{V} = (x_v, y_v)^{\mathrm{T}}$ are computed using the standard perspective projection equations $x_i = (f/(x_t - h))\, x_t$ and $y_i = (f/(x_t - h))y_t$. The vanishing point is found by taking the limit of the co-ordinates as $\lambda \to \infty$, i.e.

$$x_v = \lim_{\lambda\to\infty} x_i = \lim_{\lambda\to\infty} f\frac{a_1 + \lambda b_1}{a_3 + \lambda b_3 - h} = f\frac{b_1}{b_3},$$

$$y_v = \lim_{\lambda\to\infty} y_i = \lim_{\lambda\to\infty} f\frac{a_2 + \lambda b_2}{a_3 + \lambda b_3 - h} = f\frac{b_2}{b_3}. \tag{8}$$

Suppose that the vector $\mathbf{N} = (p, q, 1)^{\mathrm{T}}$ represents the surface-normal to the texture-plane in the viewer-centred co-ordinate system of the image. Since every line lying on the texture-plane will be perpendicular to the normal vector, then by using Eq. (8) we have

$$\mathbf{N}\cdot\mathbf{B} = pb_1 + qb_2 + b_3 = p\left(\frac{x_v b_3}{f}\right) + q\left(\frac{y_v b_3}{f}\right) + b_3 = 0. \tag{9}$$

Since $b \neq 0$, this implies that $px_v + qy_v + f = 0$. In order to solve this equation and to determine the surface plane normal vector, we need to estimate two different vanishing points in the image plane. Suppose that the two points are $\mathbf{V}_1 = (x_{v1}, y_{v1})^{\mathrm{T}}$ and $\mathbf{V}_2 = (x_{v2}, y_{v2})^{\mathrm{T}}$. The resulting normal vector components p and q are found by solving the system of simultaneous linear equations

$$\begin{bmatrix} x_{v1} & y_{v1} \\ x_{v2} & y_{v2} \end{bmatrix}\begin{bmatrix} p \\ q \end{bmatrix} = -\begin{bmatrix} f \\ f \end{bmatrix}. \tag{10}$$

The solution parameters, p and q are

$$p = f\frac{y_{v1} - y_{v2}}{x_{v1}y_{v2} - x_{v2}y_{v1}}, \qquad q = f\frac{x_{v2} - x_{v1}}{x_{v1}y_{v2} - x_{v2}y_{v1}}. \tag{11}$$

Using the two slope parameters, the slant and tilt angles are computed using the formulae

$$\sigma = \arccos\left(\frac{1}{\sqrt{p^2 + q^2 + 1}}\right), \qquad r = \arctan\left(\frac{q}{p}\right). \tag{12}$$

4. Projective distortion of the power spectrum

In Section 2 we described the affine approximation of the perspective projection using a first-order Taylor series. The approximation allows us to model the global perspective geometry using local affine patches which are quilted together. The model is similar to the scaled orthographic projection [19]. Furthermore, in Section 3 we presented the geometric relationships necessary to recover planar surface orientation from two vanishing points in the image. In this section we will develop these ideas one step further by showing how the spectral content of the locally affine patches relates to the global parameters of perspective geometry. Specifically, we will describe how the vanishing point position $\mathbf{V} = (x_v, y_v)$ can be estimated from local spectral information.

The Fourier transform provides a representation of the spatial frequency distribution of a signal. In this section we show how local spectral distortion resulting from our linear approximation of the perspective projection of a texture patch can be computed using an affine transformation of the Fourier representation. We will commence by using an affine transform property of the Fourier domain [17]. This property relates the linear effect of an affine transformation A in the spatial domain to the frequency domain distribution. Suppose that $G(.)$ represents the Fourier transform of a signal. 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., 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

$$G(\mathbf{U}_i) = \frac{1}{|\det(A)|}\, \mathrm{e}^{2\pi j\mathbf{U}_i^{\mathrm{T}}A^{-1}B} G[(A^{\mathrm{T}})^{-1}\mathbf{U}_t]. \tag{13}$$

In our case, the affine transformation is T_p^* as given in Eq. (5) and there are no translation coefficients, i.e., $\mathbf{B} = 0$. As a result Eq. (13) simplifies to

$$G(\mathbf{U}_i) = \frac{1}{|\det(T_p^*)|}\, G[(T_p^{*\mathrm{T}})^{-1}\mathbf{U}_t]. \tag{14}$$

In other words, the effect of the affine transformation of co-ordinates T_p^* induces an affine transformation $(T^{*\mathrm{T}}_p)^{-1}$ on the texture-plane frequency distribution. 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 over the frequency peaks, i.e., the energy amplitude will not be considered in the analysis. For practical purposes we will use local power spectrum as the spectral representation of the image. This describes the energy distribution of the signal as a function of its frequency content. In this way we will ignore complications introduced by phase information. Using the power spectrum, small changes in phase due to translation will not affect the spectral information and Eq. (14) will hold.

Our overall goal is to consider the effect of perspective transformation on the power-spectrum. In practice, however, we will be concerned with periodic textures in which the power spectrum is strongly peaked. In this case we can confine our attention to the way in which the dominant frequency components transform. According to our affine approximation and Eq. (14), the way the Fourier domain transforms locally is governed by

$$\mathbf{U}_i = (T_p^{*\mathrm{T}})^{-1}\mathbf{U}_t. \tag{15}$$

In the next section we will use this local spectral distortion model to establish some properties of the projected spectral distribution in the viewer-centred co-ordinate system. In particular, we will show that lines radiating from the vanishing point connect points with identically oriented spectral components. We will exploit this property to develop a geometric algorithm for recovering the image-plane position of the vanishing-point, and hence, for estimating the orientation of the texture-plane.

5. Lines of constant spectral orientation

In this section we focus on the directional properties of the local spectrum distribution. We will show how the uniformity of the spectral angle over the image plane can be used to estimate the vanishing point location and hence compute planar surface orientation.

On the texture plane the frequency-domain angle of the unprojected spectral component is given by $\beta = \arctan(v_t/u_t)$. Using the affine transformation of frequencies given in Eq. (15), and the transformation in Eq. (5), after perspective projection, the corresponding frequency domain angle in the image plane is

$$\alpha = \arctan\left[\frac{v_i}{u_i}\right] = \arctan\left[\frac{u_t f \sin\tau + v_t\,(xo_i \sin\sigma + f\cos\tau\cos\sigma)}{u_t f\cos\tau - v_t\,(yo_i\sin\sigma + f\sin\tau\cos\sigma)}\right]. \tag{16}$$

For simplicity, we confine our attention to a rotated system of image-plane co-ordinates in which the x-axis is aligned in the tilt direction. In this rotated system of co-ordinates, the above expression for the image-plane spectral angle simplifies to

$$\alpha = -\arctan\left[\frac{(f\cos\sigma + xo_i\sin\sigma)v_t}{yo_i\,v_t\sin\sigma - fu_t}\right]. \tag{17}$$

Let us now consider a line in the image plane radiating from a vanishing point which results from the projection to a family of horizontal parallel lines on the texture plane. This family of parallel lines would originally be described by the spectral component $\mathbf{U}_s = (0, v_t)^{\mathrm{T}}$. After perspective projection this family of parallel lines can be written in the "normal-distance" representation as

$$\mathscr{L} : r = xo_i\cos\theta + yo_i\sin\theta \quad \forall(xo_i, yo_i) \in \mathscr{L}, \tag{18}$$

where r is the length of the normal from the line to the origin, and, θ is the angle subtended between the line-normal and the x-axis. Fig. 4 illustrates the geometry of this line representation together with the angle α of the spectral component at the point of expansion (xo_i, yo_i). Since this line passes through the vanishing point $\mathbf{V} = (x_v, y_v) = (-f\cos\sigma/\sin\sigma, 0)^{\mathrm{T}}$ on the image plane,

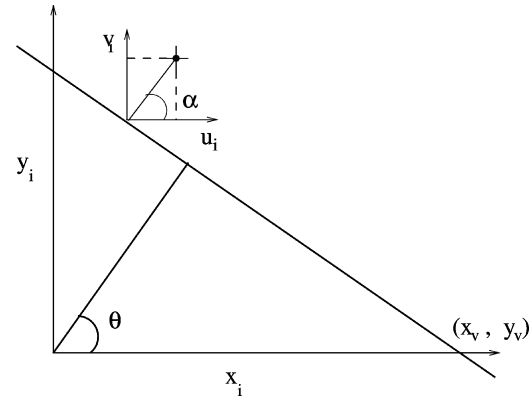

Fig. 4. Normal form for lines radiating from a vanishing point (x_v, y_v). A local spectral component centred at a point (xo_i, yo_i) point over the line is also shown.

for $f < 0$, r can also be written as

$$r = x_v\cos\theta + y_v\sin\theta = \frac{-f\cos\sigma}{\sin\sigma}\cos\theta. \tag{19}$$

Substituting Eq. (19) into Eq. (18) and solving for yo_i we obtain

$$yo_i = -\left(\frac{f\cos\sigma + xo_i\sin\sigma}{\tan\theta\sin\sigma}\right). \tag{20}$$

By substituting the above expression in Eq. (17) and using the fact that $u_s = 0$, after some simplification we find that $\alpha = \theta, \forall(xo_i, yo_i) \in \mathscr{L}$. As a result, each line belonging to the family $\mathscr{L}$ connects points on the image plane whose local spectral distributions have a uniform spectral angle α. These lines will intercept at a unique point which is a vanishing point on the image plane.

By using this property we can find the co-ordinates of vanishing points on the image plane by connecting those points which have corresponding spectral components with identical spectral angles. We meet this goal by searching lines for which the angular correlation between the spectral moments is maximum. To proceed we adopt a polar representation for the power spectrum. Suppose $P_{\eta\alpha}(\eta, \alpha)$ is the power spectrum in polar co-ordinates where $\eta = \sqrt{u_i^2 + v_i^2}$ is the radial variable and $\alpha = \arctan(v_i/u_i)$ is the angular variable. Integrating over the radial variable, the angular power-spectrum is given by

$$P_\alpha(\alpha) = \int_\eta^0 P_{\eta\alpha}(\eta, \alpha)\, d\eta. \tag{21}$$

The angular distribution of spectral power at any given image can now be matched against those of similar orientation by maximising angular correlation. For the

purpose of matching we use the normalised correlation

$$\rho = \frac{\int_\alpha P_\alpha(\alpha)P'_\alpha(\alpha)\,\mathrm{d}\alpha}{\int_\alpha P_\alpha(\alpha)\,\mathrm{d}\alpha \int_\alpha P'_\alpha(\alpha)\,\mathrm{d}\alpha}, \tag{22}$$

where $P_\alpha(\alpha)$ and $P'_\alpha(\alpha)$ are the two angular distributions being compared. The points with the highest values of ρ can be now connected to determine a line pointing in the direction of the vanishing point.

To compute the angular power distribution $P_\alpha(\alpha)$ we require a way of sampling the local power spectrum. In particular, we need a sampling procedure which provides a means of recovering the angular orientation information residing in the peaks of the power spectrum. We accomplish this by simply searching for local maxima over a filtered representation of the local power spectrum. Since we are interested in the angular information rather than the frequency contents of the power spectrum, we ignore the very low-requency components of the power spectrum since these mainly describe micro-texture patterns or very slow energy variation. In order to obtain a smooth spectral response we therefore use the Blackman–Tukey(BT) power-spectrum estimator which is the frequency response of the windowed autocorrelation function. We employ a triangular smoothing window $w(\mathbf{X})$ [20] due to its stable spectral response. The spectral estimator is then

$$P(\mathbf{U}_i)^{BT} = \mathscr{F}\{c_{xx}(\mathbf{X}_i) \times w(\mathbf{X}_i)\}, \tag{23}$$

where c_{xx} is the estimated autocorrelation function of the image patch. Providing that we have at least two representative spectral peaks we can directly generate line directions according to the angular property we have described in previous the section of this paper. We can use as many distinct spectral components as we can estimate. However, a two-component decomposition is sufficient for our purposes.

6. Experiments

In this section we provide some results which illustrate the accuracy of planar pose estimation achievable with our new 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 [21]. 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 compare the sensitivity of the new method based on multiple vanishing points with that of a gradient based or single vanishing point method.

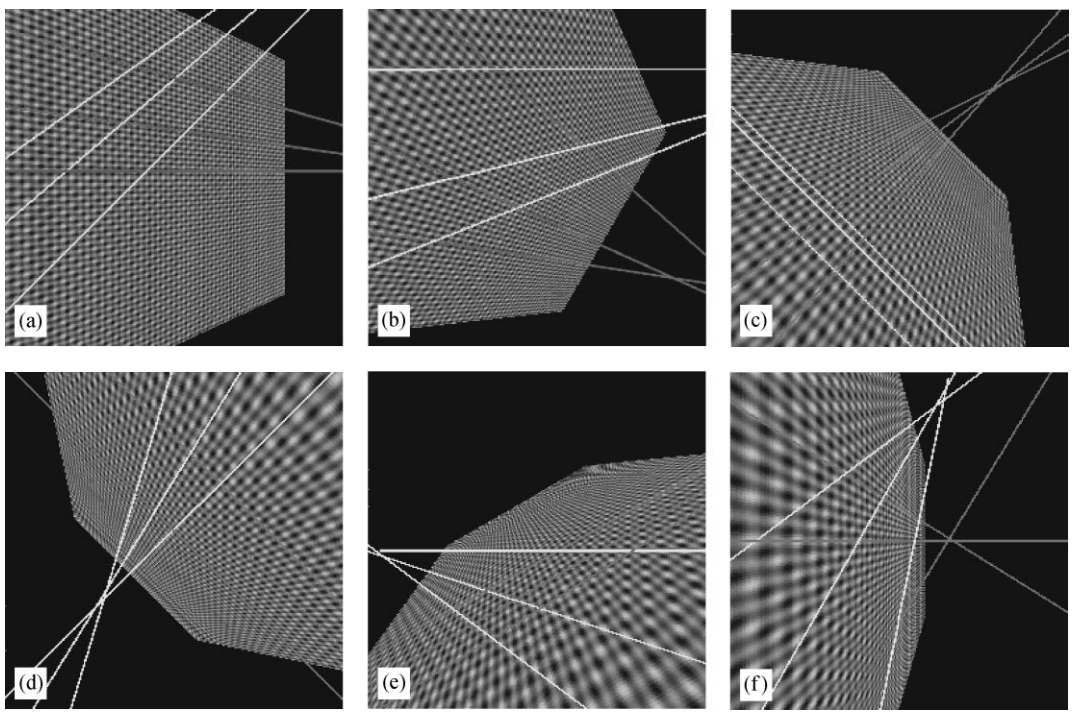

Fig. 5. Artificial texture images: Group 1—sinusoidal texture images.

6.1. Synthetic texture planes

We commence with some examples for a synthetic regular sinusoid texture. Fig. 5 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 1 we list the ground-truth and estimated orientation angles. Also listed is the absolute error. The agreement between the ground-truth and estimated values is generally very good. A point to note from these first set of images is that the estimated slant error is proportional to the slant angle. This is due to the significant variations of the texture scale for larger slant angles. This suggests that an adaptive scale process should be used to improve the accuracy of the method. From the results obtained from this set of images, there is no systematic variation in the tilt angle errors which have an average value of 0.8°. The average error for the slant angle is 0.9°.

Table 1
Actual × estimated slant and tilt values (Artificial Group 1)

Image	Actual		Estimated		Abs. Error	
	(σ)	(τ)	(σ')	(τ')	σ'	τ'
(a)	20	0	19.7	0.0	0.3	0.0
(b)	30	− 30	30.0	− 28.2	0.0	1.8
(c)	45	45	44.7	46.0	0.3	1.0
(d)	50	225	51.1	225.4	1.1	0.4
(e)	60	120	58.7	118.3	1.3	1.7
(f)	70	0	72.5	− 0.3	2.5	0.3

A second group of artificial images is shown in Fig. 6. In this group we have taken three different artificial textures and have projected them onto planes of known slant and tilt. The textures are composed of regularly spaced geometric primitives of uniform size. Specifically, we have chosen elliptical, rectangular and lattice shaped primitives. However, it is important to stress that in this case the texture elements are not oriented in the direction of the vanishing point. Superimposed on the projected textures are the estimated lines of uniform spectral orientation. In Table 2 we list the ground-truth and estimated values of the slant and tilt angles together with the corresponding errors. The agreement between the estimated and ground-truth angles is good. Moreover, the computed errors are largely independent of the slant angle. The average slant-error is 1° and the average tilt-error is 2.5°.

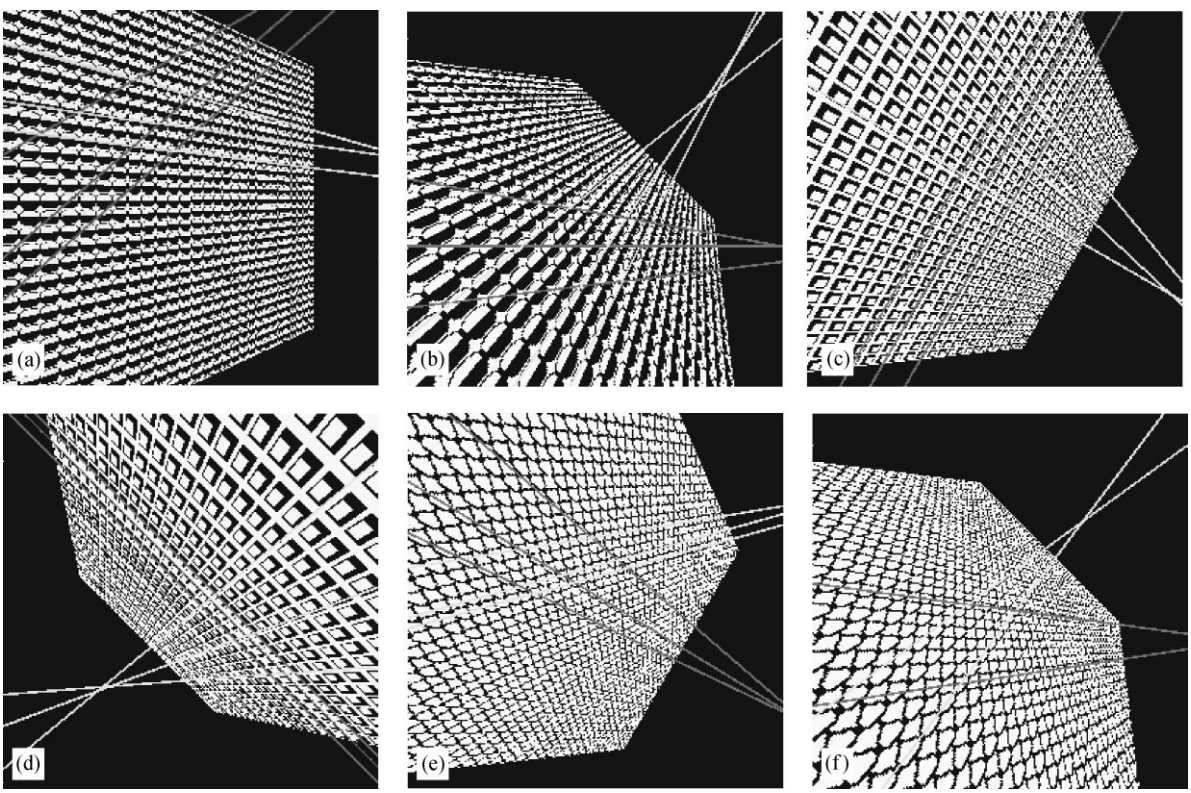

Fig. 6. Artificial texture images: Group 2—regular geometric images.

A third group of images is shown in Fig. 7. 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 textures are regular natural textures of almost regular element distribution. Superimposed on the projected textures are the estimated lines of uniform spectral orientation. The values for the estimated orientation angles are listed in Table 3.

6.2. Real-world examples

This part of the experimental work focuses on real-world textures with unknown ground-truth. The textures used in this study are two views of a brick wall, a York pantile roof and the lattice casing enclosing a PC monitor. The images were collected using a Kodak DC210 digital camera and are shown in Fig. 8. There is some geometric distortion of the images due to camera optics. This can be seen by placing a ruler or straight edge on the

Table 2
Actual × estimated slant and tilt values (Artificial Group 2)

Image	Actual		Estimated		Abs. Error	
	(σ)	(τ)	(σ')	(τ')	σ'	τ'
(a)	20	0	20.1	1.0	0.1	1.0
(b)	45	45	46.2	46.4	1.2	1.4
(c)	30	− 30	30.0	− 32.4	0.0	2.4
(d)	50	225	49.5	222.4	0.5	2.6
(e)	30	− 30	30.1	− 35.2	0.1	5.2
(f)	45	45	41.2	44.0	3.8	1.0

Table 3
Actual × estimated slant and tilt values (Brodatz Textures)

Image	Actual		Estimated		Abs. Error	
	(σ)	(τ)	(σ')	(τ')	σ'	τ'
(a)	30	0	34.5	0.0	4.5	0.0
(b)	50	225	53.9	223.5	3.9	1.5
(c)	30	0	27.5	0.0	2.5	0.0
(d)	45	45	51.7	44.6	6.7	0.4
(e)	30	− 30	30.4	− 23.3	0.4	6.7
(f)	60	120	59.6	125.0	0.4	5.0

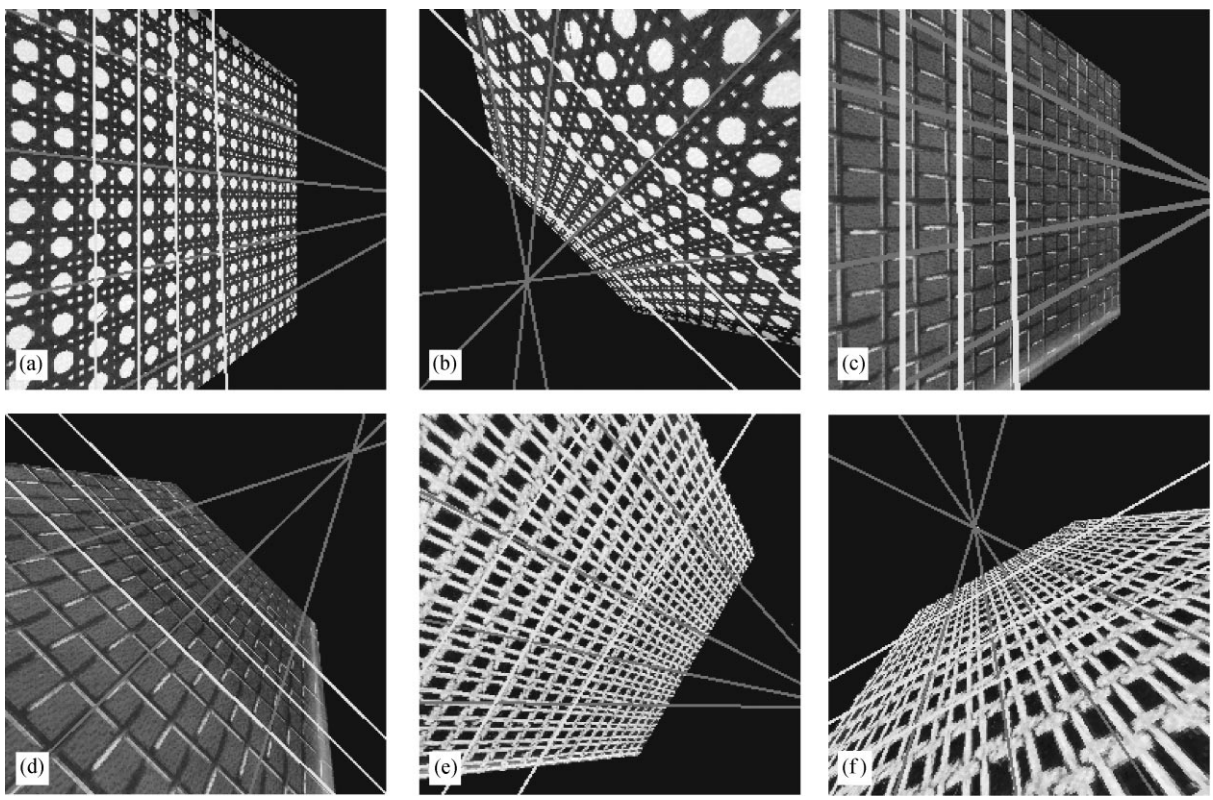

Fig. 7. Natural texture images: Group 3—Artificially projected Brodatz textures. (a) and (b) D101; (b) and (c) D1; (d) and (e) D20.

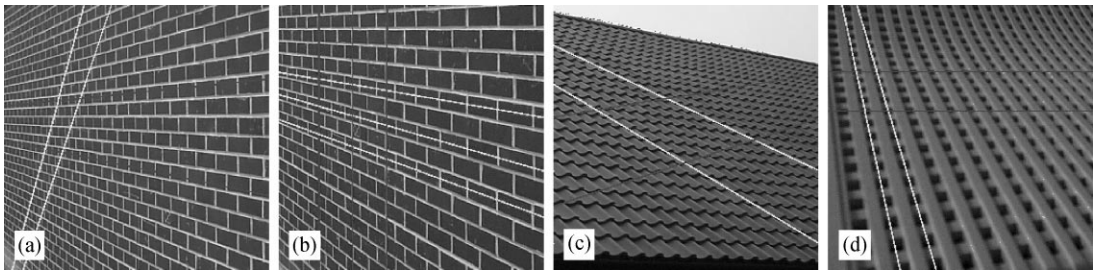

Fig. 8. Outdoor texture images. (a) and (b) Brick wall; (c) Roof; (d) PC casing.

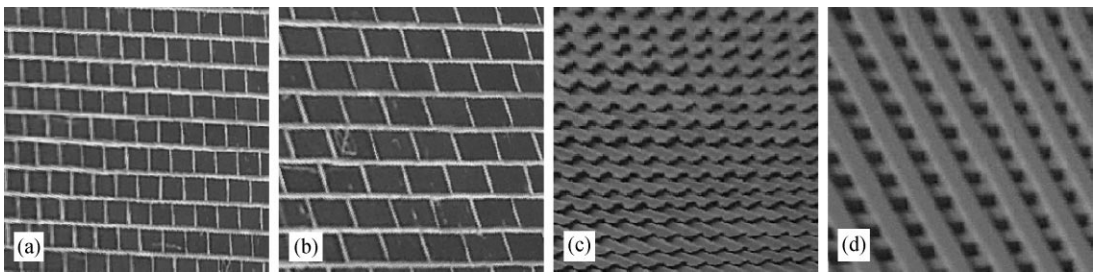

Fig. 9. Back-projected outdoor texture images. (a) and (b) Brick wall; (c) Roof; (d) PC casing.

brick-wall images and observing the deviations along the lines of mortar between the bricks.

Superimposed on the images are the lines of uniform spectral orientation. In the case of the brick-wall images these closely follow the mortar lines. In Fig. 9 we show the back-projection of the textures onto the fronto-parallel plane using the estimated orientation angles. In the case of the brick wall, any residual skew is due to error in the estimation of the slant and tilt parameters. It is clear that the slant and tilt estimates are accurate but that there is some residual skew due to poor tilt estimation.

6.3. *Sensitivity analysis*: *multiple vanishing point and the gradient based-method*

Finally, we provide some comparison between our two vanishing point method and the use of a single vanishing point in conjunction with texture gradient information [22]. 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. 5.

We commence by studying the errors for the slant and tilt angle estimation when only the slant angle varies. The graphs in Fig. 10 plot the slant and tilt errors when the orientation varies from 10° to 80° and the tilt angle remains constant at 0°. In Fig. 10(a), the slant error is considerably larger for the gradient-based method than for the multiple vanishing point method. Moreover, our new method provides better estimates for small slant angles. It is worth commenting that since we operate with a single fixed-scale representation of the power spectrum, there is a danger of texture sampling errors for large slant angles. However, the average slant error is still small and around 2°. The average slant error for the gradient method is this case is 9.5°.

Turning our attention for the tilt error in Fig. 10(b), we observe that the new method accurately estimates the tilt angle, except when the slant angle is greater than 50°. Again the fixed-scale problem can be observed. The accuracy of the gradient-based method is low for very small slant angles. This is due to difficulties in estimating the direction of the energy gradient due to small variations in texture density with location on the image plane. In the case of tilt estimation, both methods have similar accuracy for medium slant angles. However, the gradient-based method also looses accuracy for larger slant angles due to problems related to the texture scale.

In Fig. 11 we show the estimation error under varying tilt and fixed slant angle. The tilt varies over the range from 0° to 170°, whilst the slant angle remains constant at 40°. In Fig. 11(a) we show the slant error. The error for the new multiple vanishing point method is very small and almost constant with an average of 0.8°. By contrast, the average error for the gradient method is 13°. In Fig. 11(b) where we plot the tilt error, there is little to distinguish the two methods. The average value for the tilt error for the new method is 2.4° while for the gradient-based method it is 2.8°.

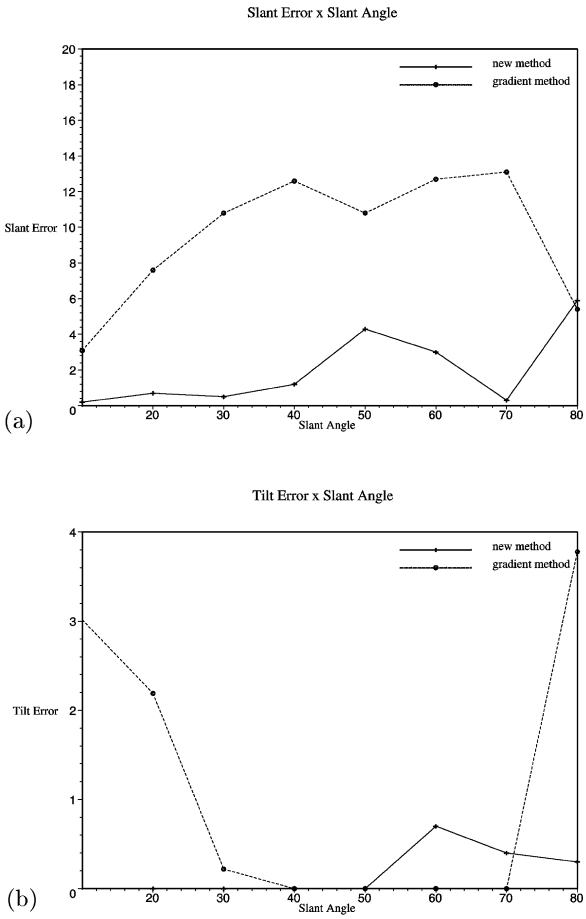

Fig. 10. Slant/tilt error plot for both methods. Image slant varying from 10 to 80°. (a) estimated slant error; (b) estimated tilt error.

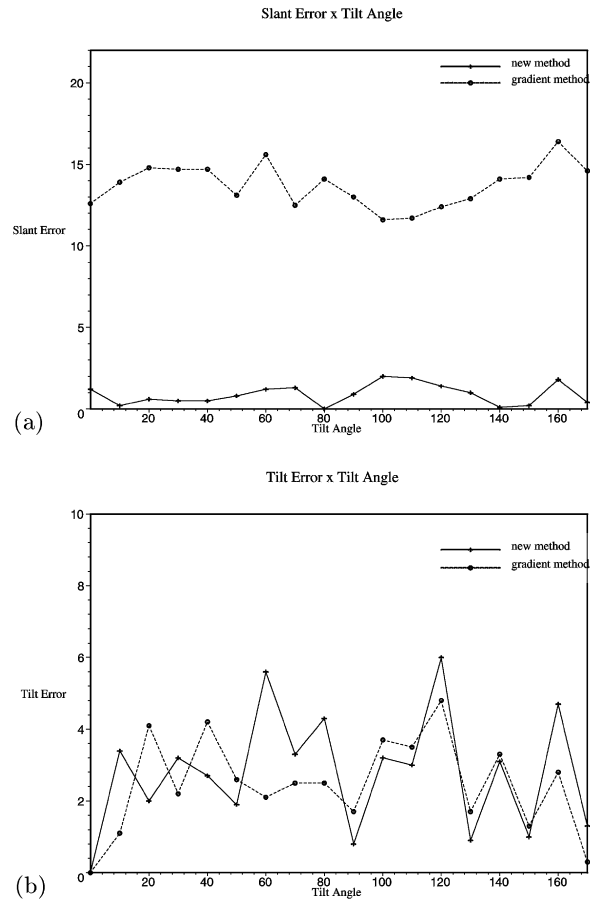

Fig. 11. Slant/tilt error plot for both methods. Image tilt varying from 0 to 170°. (a) estimated slant error; (b) estimated tilt error.

7. Conclusions

We have described an algorithm for estimating the perspective pose of textured planes from pairs of vanishing points. The method searches for sets of lines that connect points which have identically oriented spectral moments. These lines intercept at the vanishing point. The main advantage of the method is that it does not rely on potentially unreliable texture gradient estimates to constrain the tilt angle. As a result the estimated tilt angles are more accurately determined.

There are a number of ways in which the ideas presented in this paper can be extended. In the first instance, we are considering ways of improving the search for the vanishing points. Specific candidates include Hough-based voting methods. The second line of investigation is to extend our ideas 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, 1950.
[3] E.J. Cutting, R.T. Millard, Three gradients and the perception of flat and curved surfaces, J. Exp. Psychol. 113 (2) (1984) 198–216.
[4] K.A. Stevens, On texture gradients, J. Exp. Psychol. 113 (2) (1984) 217–220.
[5] J.R. Kender, Shape from texture: an aggregation transform that maps a class of texture into surface orientation, in: 6th IJCAI, Tokyo, 1979, pp. 475–480.
[6] J.S. Kwon, H.K. Hong, J.S. Choi, Obtaining a 3-d orientation of projective textures using a morphological method, Pattern Recognition 29 (1996) 725–732.

[7] K. Ikeuchi, Shape from regular patterns, Artif. Intell. 22 (1984) 49–75.
[8] J. Aloimonos, M.J. Swain, Shape from texture , Biol. Cybernet. 58 (5) (1988) 345–360.
[9] K. Kanatani, T. Chou, Shape from texture: general principle, Artif. Intell. 38 (1989) 1–48.
[10] J. Krumm, S.A. Shafer, Texture segmentation and shape in the same image, IEEE Interational Conference on Computer Vision, 1995, pp. 121–127.
[11] B.J. Super, A.C. Bovik, Planar surface orientation from texture spatial frequencies, Pattern Recognition 28 (5) (1995) 729–743.
[12] M.J. Black, R. Rosenholtz, Robust estimation of multiple surface shapes from occluded textures, in: IEEE International Symposium on Computer Vision, 1995, pp. 485–490.
[13] 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, 1993, pp. 267–273.
[14] R. Bajcsy, L. Lieberman, Texture gradient as a depth cue, Comput. Graphics Image Process. 5 (1976) 52–67.
[15] 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.
[16] 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, pp. 353–364.
[17] 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.
[18] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, Addison-Wesley, Reading, MA, 1993.
[19] J. Aloimonos, Perspective approximations, Image Vision Compu. 8 (3) (1990) 179–192.
[20] S.M. Kay, Modern Spectral Estimation: Theory and Application, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[21] P. Brodatz, Textures: A Photographic Album for Artists and Designers, Dover, New York, 1966.
[22] 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.

About the Author—ERALDO RIBEIRO is currently undertaking research towards 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 gained his B.Sc. in physics in 1977 and Ph.D. in high energy nuclear Physics in 1981, both from the University of Durham, UK. After a period of postdoctoral research working on charm-photo-production experiments at the Stanford Linear Accelerator Centre, he moved into the fields of computer vision and pattern recognition in 1985. Between 1981 and 1991, he held posts at the Rutherford–Appleton Laboratory, the Open University and the University of Surrey. He joined the University of York as a lecturer in the Department of Computer Science in July 1991. After being promoted to Senior Lecturer in October 1997 and to Reader in October 1998, he was appointed Professor of Computer Vision in December 1998. He leads a group of some 15 researchers in the areas of computer vision and pattern recognition. He has published about 180 refereed papers in the fields of high energy nuclear physics, computer vision, image processing and pattern recognition. He was awarded the 1990 Pattern Recognition Society Medal and received an Outstanding Paper Award in 1997. Professor Hancock serves as an Associate Editor of the journal Pattern Recognition and has been a guest editor for the Image and Vision Computing Journal. He is currently guest-editing a special edition of the Pattern Recognition journal devoted to energy minimisation methods in computer vision and pattern recognition. He chaired the 1994 British Machine Vision Conference and has been a programme committee member for several national and international conferences.