Accumulating Spectral Evidence for Perspective Views of Texture Planes

Eraldo Ribeiro* and Edwin R. Hancock
Department of Computer Science,
University of York, York Y01 5DD, UK

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 a 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. We experiment with the new shape-from-texture technique on planar textures in buildings.

1 Introduction

The perspective foreshortening of surface patterns is an important cue for the recovery of surface orientation from 2D images [2, 7]. Broadly speaking there are two routes to recovering the parameters of perspective projection for texture patterns. The first of these is to estimate the texture gradient [3, 11]. 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 [8, 9, 16]. Provided that two or more vanishing points are available, then planar surface orientation can be directly determined.

Unfortunately the location of 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 [8, 9]. Moreover, multiple vanishing point detection may

*Supported by CAPES-BRAZIL, under grant: BEX1549/95-2

BMVC 1999 doi:10.5244/C.13.26

be even more elusive. It is for this reason that frequency domain analysis offers an attractive alternative [5, 1, 6]. 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.

In this paper we make use of an interesting spectral property which provides a direct route to vanishing point location via the use of frequency domain information. The observation is a simple one. At each point on the image plane, the spectral angle points in the direction of a vanishing point. Lines that radiate from a vanishing point therefore connect points of uniform spectral angle.

Although this spectral property can be used to directly locate vanishing point positions on the image plane, and hence estimate perspective pose, there are a number of implementational difficulties. Firstly, the vanishing point may fall anywhere on the extended image plane. In other words, the search space is not bounded. Secondly, in order to triangulate the vanishing points, several lines must be fitted through points which have identical spectral orientation. As a result, spectral correspondences must be established in order to fit the lines. The aim in this paper is to provide a representation which provides a single, and quite elegant, solution to these two problems.

The idea underpinning our method is to exploit the unit-sphere representation of the image plane [16]. This involves placing a sphere of unit radius at the focal point of the camera. Lines on the image planes map to great circles on the unit sphere. This representation bounds the search space for vanishing points, since parallel lines meet at opposite poles of the unit sphere. This solves our first problem of bounded search. Our solution to the problem of finding correspondences is to treat the unit-sphere as an accumulator space. The position of a point on the image plane together with the direction of its spectral moment define a line. Each such line transforms to a great circle on the unit sphere. When several local spectra radiate from the same vanishing point on the image plane, then their great circles will intersect at a common location on the unit sphere. By accumulating votes along great circles on a suitably quantised representation of the unit-sphere we can search for vanishing points. These correspond to local maxima of accumulation.

2 Perspective Modelling

We commence by reviewing the projective geometry for the perspective transformation of points on a plane [5, 1]. 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. 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 $\mathbf{X_t} = (x_t, y_t)^T$ on the texture plane and $\mathbf{X_i} = (x_i, y_i)^T$ on the image plane. The perspective transformation between the two co-ordinate systems is

$$\mathbf{X_i} = T_p \mathbf{X_t} \tag{1}$$

We represent the orientation of the viewed surface plane using the slant σ and tilt τ angles. This parametrisation is a natural way to model local surface orientation.

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 around the line of sight 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 Equation 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. The first-order Taylor approximation to the transformation is:

$$T_p^* = \frac{\Omega}{hf\cos\sigma} \begin{bmatrix} xo_i \sin\sigma + f\cos\tau\cos\sigma & -f\sin\tau \\ yo_i \sin\sigma + f\sin\tau\cos\sigma & f\cos\tau \end{bmatrix} \tag{3}$$

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 Equation 3 operates from the texture plane to the image plane. The model is similar to the scaled orthographic projection [4].

The net effect of the global perspective transformation is to distort the viewer-centred texture pattern in the direction of vanishing points on the image plane. Given two points $\mathbf{V_1} = (x_{v1}, y_{v1})^T$ and $\mathbf{V_2} = (x_{v2}, y_{v2})^T$ we can directly determine the 3-D orientation of the plane. Let the normal-vector to the texture plane be $N = (p, q, 1)$. The resulting normal vector components p and q are found by solving the equations [13]:

$$p = f\frac{y_{v1} - y_{v2}}{x_{v1}y_{v2} - x_{v2}y_{v1}} \qquad q = f\frac{x_{v2} - x_{v1}}{x_{v1}y_{v2} - x_{v2}y_{v1}} \tag{4}$$

Using the two slope parameters, the slant and tilt angles are computed using $\sigma = \arccos\left(\frac{1}{\sqrt{p^2+q^2+1}}\right)$ and $\tau = \arctan\left(\frac{q}{p}\right)$. If more than two vanishing points are available, then the recovery of perspective pose parameters is over-constrained and can be effected by least-squares estimation.

3 Projective Distortion of the Power Spectrum

The Fourier transform provides a representation of the spatial frequency distribution of a signal. The novel contribution 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 [14]. This property relates the linear effect of an affine transformation A in the spatial domain to the frequency domain distribution. Suppose that $F(.)$ 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* [14], 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

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

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

$$F(\mathbf{U_i}) = \frac{1}{|det(T_p^*)|} F[T_p^{*-T}\mathbf{U_t}] \tag{6}$$

In other words, the effect of the affine transformation of co-ordinates T_p^* induces an affine transformation T_p^{*-T} 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 the local power spectrum as the spectral representation of the image. This describes the energy distribution of the image 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 hence Equation 6 will hold. The power spectrum representation of an image $f(\mathbf{X_t})$ may be defined as the Fourier transform of the autocorrelation function of the image.

In order to obtain a smooth spectral response we use the Blackman-Tukey(BT) power spectrum estimator. This is the frequency response of the windowed auto-correlation function. We employ a triangular smoothing window $w(\mathbf{X})$ [15] due to its stable spectral response. The spectral estimator is then

$$P(\mathbf{U_i})^{BT} = \mathcal{F}\{c_{xx}(\mathbf{X_i}) \times w(\mathbf{X_i})\} \tag{7}$$

Where c_{xx} is the estimated autocorrelation function of the image patch. Our overall goal is to consider the effect of perspective transformation on the power-spectrum. In practice, however, we will be concerned with semi-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 Equation 6, the way the Fourier domain transforms locally is governed by

$$\mathbf{U_i} = T_p^{*-T}\mathbf{U_t} \tag{8}$$

This spectral property has also been exploited by Rosenholtz and Malik [6] in their work on local shape-from-texture.

4 Lines of Constant Spectral Orientation

We will now consider the angular distortion of the local spectral distribution resulting from perspective projection. On the texture-plane the frequency-domain

angle of the unprojected spectral component is given by $\beta = \arctan\left(\frac{v_t}{u_t}\right)$. Using the affine transformation of frequencies given in Equation 8 and the inverse transpose of the perspective transformation in Equation 3, the corresponding frequency domain angle in the image plane is

$$\alpha = \arctan\left[\frac{v_i}{u_i}\right] = \arctan\left[\frac{u_t f \sin\tau + v_t\left(xo_i \sin\sigma + f\cos\tau\cos\sigma\right)}{u_t f\cos\tau - v_t\left(yo_i\sin\sigma + f\sin\tau\cos\sigma\right)}\right] \tag{9}$$

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

$$\alpha = -\arctan\left[\frac{(f\cos\sigma + xo_i\sin\sigma)v_t}{yo_i v_t\sin\sigma - f u_t}\right] \tag{10}$$

Let us now assume that the line $\mathcal{L}_\theta$ radiates from a vanishing point which results from the projection to a family of horizontal parallel lines on the texture plane. Before projection, this family of parallel lines is described by the spectral component $\mathbf{U_s} = (0, v_t)^T$. After perspective projection, this family of lines can be written in the "normal-distance" representation as

$$\mathcal{L}_\theta : r_\theta = xo_i\cos\theta + yo_i\sin\theta \qquad \forall\,(xo_i, yo_i) \in \mathcal{L}_\theta \tag{11}$$

where r_θ 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_i-axis. Figure 1 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). Here we are particularly interested in the angle of the local spectral components along one of the lines radiating from a vanishing point. If (xo_i, yo_i) belongs to such a line $\mathcal{L}_\theta$, then we can rewrite Equation 10 as:

$$\alpha = -\arctan\left[\frac{(f\cos\sigma + xo_i\sin\sigma)v_t}{(r_\theta - xo_i\cos\theta)\,v_t\sin\sigma - f u_t}\right] \tag{12}$$

Since $\mathcal{L}_\theta$ passes through the horizontal vanishing point $\mathbf{V} = (x_v, y_v) = (-\frac{f\cos\sigma}{\sin\sigma}, 0)^T$ on the image plane, for $f < 0$, the normal distance of the line r_θ is given by

$$r_\theta = x_v\cos\theta + y_v\sin\theta = \frac{-f\cos\sigma}{\sin\sigma}\cos\theta \tag{13}$$

Therefore, by substituting the above value of r_θ into Equation 12 and restricting our attention to horizontal textures for which $u_s = 0$, after some simplification we find that $\alpha = \theta$, $\forall\,(xo_i, yo_i) \in \mathcal{L}_\theta$. As a result, each line belonging to the family $\mathcal{L}_\theta$ connects points on the image plane whose the local spectral distributions have a uniform spectral angle α. Furthermore, the image plane position and the associated spectral angle estimate are sufficient to specify the equation of a putative line radiating from the vanishing point. Several such lines will intercept at a unique vanishing point on the image plane. This is a novel observation which is pivotal to the development of our new method for recovering perspective pose.

Finally, it is important to stress that our method is only limited by the need for spectra with distinct and sharply focussed components. 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 [16] and model the image plane in terms of spherical coordinates 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 the two angles of azimuth and zenith or elevation. Spherical projections of the image plane have been exploited by several authors [10, 16]. However, they have employed structural representations of texture. Instead, we use the local spectral frequency to model texture.

Figure 1 illustrates the projection geometry. The unit sphere is placed at the focal point and the image plane lies at a distance f along the optical axis. For each point on the image plane, the position (x_i, y_i) and the measured spectral angle α_i specify the equation of a line. The orientation $\theta = \alpha_i$ and the normal distance is given by $r_\theta = \sqrt{x_i^2 + y_i^2}\cos(\phi_i - \alpha_i)$ (where $\phi_i = \arctan\frac{y_i}{x_i}$). Each such line, $\mathcal{L}_\theta$

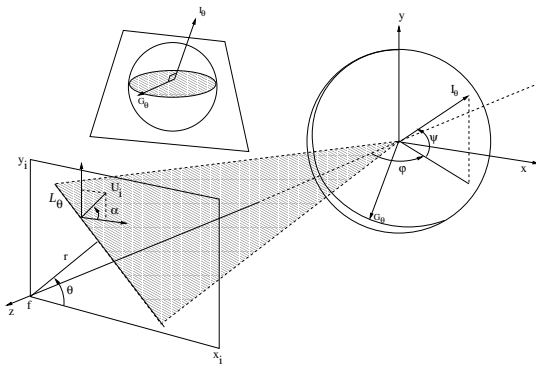


Figure 1: Projecting lines from the image plane onto the unit-sphere.

projects onto a great circle on the unit sphere. The great circles are constructed by intersecting the sphere by the plane that both contains the line $\mathcal{L}_\theta$ and the center of the sphere. Let $\mathbf{G}_\theta$ be a vector that points from the centre of the unit sphere to a point on the corresponding great circle. If the point has azimuth φ and elevation ψ, then the vector is given by $\mathbf{G}_\theta = (\sin\varphi\cos\psi, \sin\psi, \cos\varphi\cos\psi)$. Suppose that I_θ is the normal vector of the plane which contains the great circle. Figure 1 shows in the geometry of the plane normal vector I_θ and the great circle generator vector G_θ. Since the two vectors are perpendicular to one-another they satisfy the condition that $\mathbf{G}_\theta.\mathbf{I}_\theta = 0$. The azimuth and elevation angles of points on the great-circle are related to the normal distance parameters of the straight-line in the following manner

$$f\,(\sin\theta.\tan\psi + \cos\theta\sin\varphi) = r_\theta\cos\varphi \tag{14}$$

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 onto the unit-sphere:

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

We now exploit Property 1 to directly relate the local spectral angle to great circles on the unit sphere. Using the equality between the angles θ and α and using the expression for a great circle in Equation 14, we find

$$\psi = -\arctan \frac{f \cos\alpha \sin\varphi - r_\theta \cos\varphi}{f \sin\alpha} \tag{15}$$

Property 2 (From Spectral Frequency Angles to Great Circles). Each spectral angle α 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.

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 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 frequency components of the power-spectrum since these mainly describe micro-texture patterns or very slow energy variations. Providing that we have at least two representative spectral peaks we can directly generate line directions according to the angular constancy property. We can use as many distinct spectral components as we can estimate. 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. Using Equation 15 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. Once two or more intersection points are located, then the perspective pose of the plane can then be determined as described in Section 2.

7 Experiments

In this section we provide some results which illustrate the accuracy of planar pose estimation achievablewith our shape-from-texture algorithm. This evaluation is divided into two parts. We commence by considering textures with known ground-truth slant and tilt. This part of the study is based on projected Brodatz textures [12]. The second part of our experimental study focuses on real texture planes where the ground truth is unknown.

7.1 Synthetic texture planes

In Figure 2, we have taken three different texture images from the Brodatz album and haveprojected them onto planes of known slant and tilt. The textures are regular real 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.Figure 3 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 1. The main feature to note is that the method performs well even when the texture plane is highly inclined.

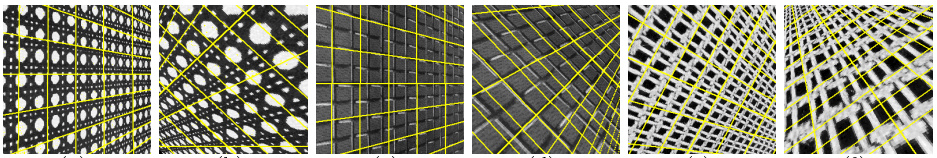


Figure 2: *Brodatz textures. (a)-(b) D101; (b)-(c) D1; (d)-(e) D20.*

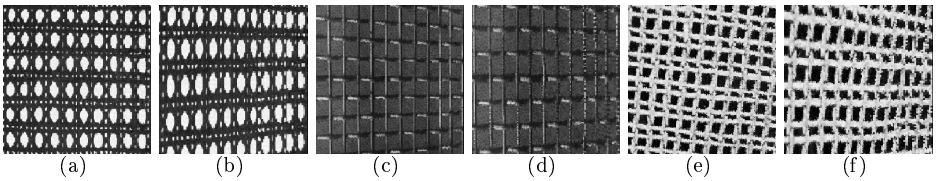


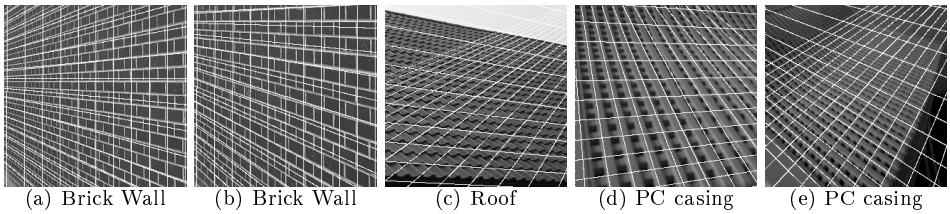
Figure 3: *Back-projected Brodatz textures. (a)-(b) D101; (b)-(c) D1; (d)-(e) D20.*

TABLE 1 - Actual× Estimated slant and tilt values (Brodatz Textures)

	actual		estimated		abs. error	
	(σ)	(τ)	(σ')	(τ')	σ'	τ'
(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

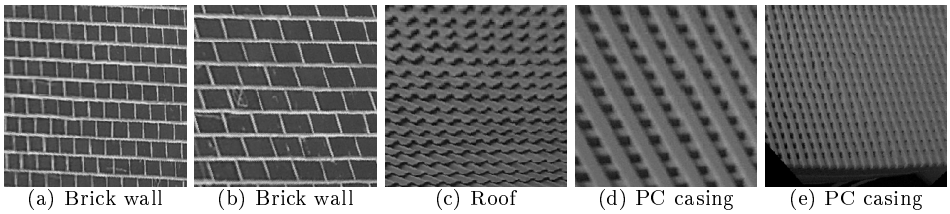
7.2 Real World Examples

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 Figure 4. 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 brick-wall images and observing the deviations along the lines of mortar between the bricks.



(a) Brick Wall (b) Brick Wall (c) Roof (d) PC casing (e) PC casing

Figure 4: *Outdoor texture images.*



(a) Brick wall (b) Brick wall (c) Roof (d) PC casing (e) PC casing

Figure 5: *Back-projected Outdoor images.*

Superimposed on the images are the lines radiating from the vanishing points. In the case of the brick-wall images these closely follow the mortar lines. In Figure 5 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 estimates are accurate but that there is some residual skew due to poor tilt estimation.

8 Conclusions

We have described an algorithm for estimating the perspective pose of textured planes by projecting spectral information onto a 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.

References

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

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

[3] Cutting E.J. and Millard R.T. Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology:general*, 113(2):198–216, 1984.

[4] Aloimonos J. Perspective approximations. *Image and Vision Computing*, 8(3):179–192, August 1990.

[5] Krumm J. and Shafer S.A. Shape from periodic texture using spectrogram. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 284–289, 1992.

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

[7] Gibson J. J. *The perception of the visual world*. Houghton Miffin, Boston, 1950.

[8] Kender J.R. Shape from texture: an aggregation transform that maps a class of texture in to surface orientation. In *6th IJCAI, Tokyo*, pages 475–480, 1979.

[9] Kwon J.S., Hong H.K., and Choi J.S. Obtaining a 3-d orientation of projective textures using a morphological method. *Pattern Recognition*, 29:725–732, 1996.

[10] Ikeuchi K. Shape from regular patterns. *Artificial Intelligence*, 22:49–75, 1984.

[11] Stevens K.A. On gradients texture gradients. *Journal of Experimental Psychology:general*, 113(2):217–220, 1984.

[12] Brodatz P. *Textures:A Photographic Album for Artists and Designers*. Dover, New York, 1966.

[13] Haralick R.M. and L.G. Shapiro. *Computer and Robot Vision*. Addison-Wesley, Reading-Massachusetts, 1993.

[14] Bracewell R.N., Chang K.-Y., Jha A.K., and Wang Y.-H. Affine theorem for two-dimensional fourier transform. *Electronics Letters*, 29(3):304, 1993.

[15] Kay S.M. *Modern Spectral Estimation:Theory and Application*. Prentice Hall, 1988.

[16] Barnard S.T. Interpreting perspective images. *Artificial Intelligence*, 21:435–462, 1983.