Comparison of Spatial-Frequency Shape-from-Texture Methods

David Mould

Supervisor: Herb Yang

Abstract

One important source of information for surface recovery is the textures on those surfaces. Textures may be viewed as periodic signals, leading to frequency-based approaches to the problem of surface recovery.

Some recent spatial-frequency methods are reviewed and their behaviour compared. Comments on the merit of the spatial-frequency approach are made.

1 Introduction

One of the central problems in computer vision is the reconstruction of three-dimensional arrangement of objects from a two-dimensional image. The texture of surfaces in the image is an important source of information about the orientation of those surfaces. *Shape-from-texture* is the problem of recovering surface orientations from textures in the image.

Many existing methods attempt to characterize textures by identifying basic texture elements or *texels*. These methods experience difficulty when no such elements are present or identifiable; not all textures can be decomposed in this way.

An alternative approach to describing a texture is to view the image data as samples of a two-dimensional signal and to compute its frequency spectrum. In order to retain spatial information, the frequency information is calculated based on local windows and each such frequency spectrum associated with the location of this window's centre.

This paper concerns itself with a particular class of shape-from-texture methods, specifically those which make use of spatial frequency information. The original approach to a spatial-frequency solution to the shape-from-texture method involved the Fourier transform; more elaborate methods have been suggested in recent years.

The main advantage of spatial-frequency shape-from-texture methods is their generality. Spatial-frequency methods reduce the required number of assumptions necessary to recover the image.

In order to make any progress on the shape recovery problem, some restrictive assumptions must be made concerning the nature of the image. Given any monochrome textured image, it is possible to conceive of a vast number of highly improbable collections of objects which would give rise to that image. In particular, texture markings can be present in such a pattern as to mimic the effects of distortion due to viewing angle.

The two most frequently used assumptions are *isotropy*, that the texture elements have no preferred direction, and *homogeneity*, that the density of texture elements does not vary. Although the second of these is satisfied by many natural textures there are many frequently-occurring textures which violate the first assumption. Figure 1 illustrates these terms with synthetic textures.

Surface orientation at a point may be described by two values [1], *slant* σ and *tilt* τ. The slant is the angle between the normal and the horizontal and the tilt is the rotational angle between this normal and an axis. Figure 2 provides an example of this relationship.

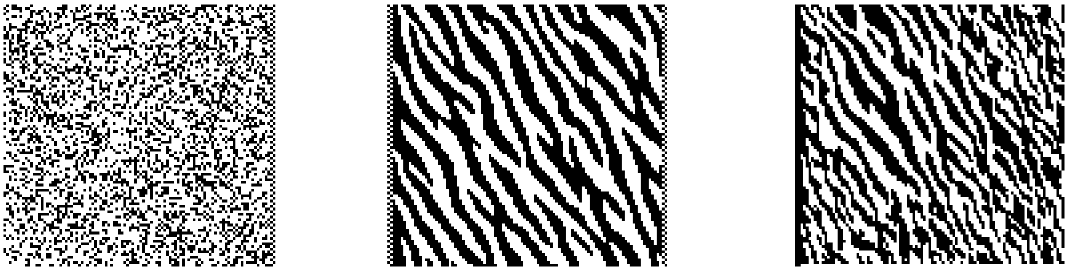

Figure 1: The leftmost texture is both isotropic and homogeneous; the centre texture is homogeneous but anisotropic; the rightmost texture is neither isotropic nor homogeneous.

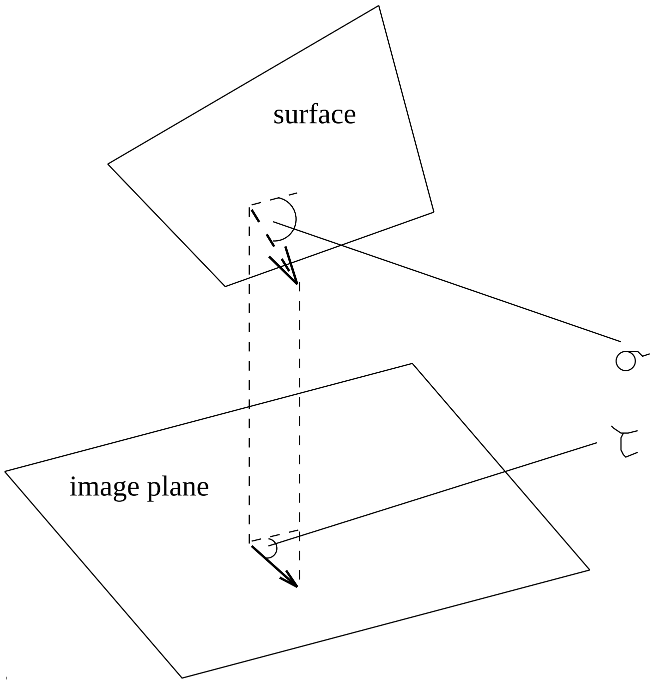

Figure 2: Diagram of slant and tilt values.

Note that the result of computing local spatial-frequency information is four-dimensional: there are two spatial dimensions and one frequency dimension for each spatial direction. For simplicity, spatial coordinates shall be (x,y) and frequency coordinates shall be (u,v).

1.1 Fourier transform

Since the spatial-frequency methods depend on obtaining a frequency spectrum for a signal, it is worthwhile to review the process of the Fourier transform and to clarify what is meant by the term *frequency spectrum.*

In brief, the Fourier transform expresses a function with period l as an infinite series:

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty}(a_m \cos\frac{m\pi x}{l} + b_m \sin\frac{m\pi x}{l})$$

The Fourier transform is the process of determining values for the coefficients a_m, b_m. This is done by exploiting the orthogonality of the sine and cosine functions:

$$\int_{-l}^{l} \cos m\pi x \cos n\pi x dx = \left\{ \begin{array}{cc} 0 & m \neq n \\ l & m = n \end{array} \right\}$$

$$\int_{-l}^{l} \cos m\pi x \sin n\pi x dx = \left\{ \begin{array}{cc} 0 & m \neq n \\ l & m = n \end{array} \right\}$$

$$\int_{-l}^{l} \sin m\pi x \sin n\pi x dx = \left\{ \begin{array}{cc} 0 & m \neq n \\ l & m = n \end{array} \right\}$$

Hence by using (2) in conjunction with (1) the following is obtained:

$$\int_{-l}^{l} f(x) \cos m\pi x dx = l a_m$$

or

$$a_m = \frac{1}{l}\int_{-l}^{l} f(x) \cos m\pi x dx$$

Similar equations hold for b_n. These are known as the Euler-Fourier formulas.

The derivation given above follows that of Boyce and DiPrima [2].

In general, a process similar to the Fourier transform can be carried out using any set of basis functions $u_n(x)$ provided that these functions are mutually orthogonal and that the set of functions is complete. The frequency spectrum is the collection of coefficients obtained by such a representation. It should be mentioned that the Euler-Fourier coefficients a_m, b_m are often conflated to $c_m = \sqrt{a_m^2 + b_m^2}$. This removes all phase information while retaining information on the frequency composition. Such a formulation is known as the spectrogram.

In an image-processing context, the Euler-Fourier formulae would be replaced with their discrete counterparts, since image data typically consists of brightness or color values at spatial intervals rather than any sort of analytic information.

Image data is finite in extent. Hence, it may be viewed as periodic with period equal to its extent without loss of generality; a Fourier or related transform is therefore meaningful. Spatial-frequency methods typically perform such transformations on some subset of the image data.

Spatial-frequency methods are useful because of the periodic nature of textures. Viewing image data in the frequency domain allows regularities to be detected and quantified. This is the basis of spatial-frequency shape-from-texture methods.

One parameter present in many spatial-frequency shape-from-texture methods is the window size. There are two opposing issues to consider: first, too small a window will be insufficient to capture the periodicity of the texture; second, too small a window will suppress the importance of spatial information. In addition, larger windows will increase the computational effort needed to produce each frequency spectrum.

2 Spectrogram

The spectrogram is the local Fourier power spectrum. The Fourier transform is computed and the frequency components c_n determined.

Krumm and Shafer [3] attempt to predict shifts in frequency due to surface shape and give a shape-from-texture algorithm based on the spectrogram. Their algorithm assumes that an orthogonally viewed texture will have spatially invariant frequency makeup.

Krumm and Shafer justify their use of the spectrogram, as opposed to other well-known methods of extracting the frequency spectrum, on the grounds that it provides visually intuitive results and is based on the well-understood Fourier transform. They note that their choice of representation does not affect the underlying theory of frequency projection.

Krumm and Shafer show that the spectrogram of the image texture is related to that of the orthogonally viewed texture by an affine transformation. Consequently, different local spectrograms are also related to one another by an affine transformation; the parameters governing this transformation are the focal length of the camera, the location of the image point, and the depth and orientation of the surface point. The complexity of this formulation is reduced by making the assumption that the surface to be recovered is planar.

The algorithm Krumm and Shafer give for recovering the surface orientation involves finding an affine transformation which best maps the spectrogram for one patch onto that of another. From the results of this computation the orientation of the plane can be determined.

The Krumm and Shafer algorithm requires that the frequency composition of the texture remain invariant across the surface. Any variation in the local frequency spectra may cause trouble for the algorithm. This is a highly restrictive assumption, since it precludes noise of any form. As the results show, even highly regular natural textures present difficulties.

The algorithm computes the local frequency spectrum at two points and performs an exhaustive search of the parameter space of the affine transformation: affine transformations for values of the two parameters evenly distributed on a grid are calculated and the closest result in a least-squares sense is considered the solution.

The computational burden of an exhaustive search is immense. Krumm and Shafer reduce the required effort in two ways: first, the search is not made over the whole of the potentially infinite parameter space; second, the values of the parameters are discretized to a high degree. The first of these places further restrictions on the domain to which the algorithm is relevant: the implementation used by Krumm and Shafer permits a maximum slant of approximately 63^o. The second introduces an unwelcome degree of granularity in the output.

An alternative to the exhaustive search is some sort of intelligent search technique. Many numerical methods to minimize functions exist; however, it may be difficult to formulate the problem in a manner amenable to solution in this way. Nevertheless, many general heuristics for search may be applied; for example, the notion of a greedy algorithm could be used here.

Intelligent search techniques assume a local coherence to the behaviour of the least-squares function in parameter space. This is a further explicit assumption about the problem. However, this assumption is implicitly present in the exhaustive search technique used: although the minimum among the points computed is recognized to differ from the actual minimum, the actual minimum is assumed to be nearby - that is, nearer to this point than to any other.

Furthermore, an intelligent search technique could use a variable step size and so avoid the granularity problems associated with the exhaustive search.

One difficulty arising from use of an intelligent search technique is that subsequent computations depend on the results of prior computations. This reduces the ease with which the algorithm can be parallelized; the present exhaustive search is trivially parallelizable, due to the absence of any need for communication during computation.

The exhaustive search method could itself be improved by changing the values at which sample calculations are performed. Rather than evenly distributing samples in parameter space, samples could be evenly distributed in σ, τ space. This would have the advantage of removing the limitation on allowed σ with little additional computational effort.

Krumm and Shafer claim that their algorithm is able to deal with aliasing. This is done with an extension to the algorithm which aliases the frequencies of the second patch according to

$$(u_{alias}, v_{alias}) = (u - \lfloor u + \frac{1}{2} \rfloor, v - \lfloor v + \frac{1}{2} \rfloor)$$

The assumption that the first patch is not aliased places serious restrictions on the circumstances in which this approach will be successful. If aliasing is a genuine problem, it is unlikely that an unaliased region can easily be found, even provided that one exists. In particular, if the method is to be extended to be entirely automated, the problem of finding an unaliased region must be investigated. Recognizing the absence of aliasing is a nontrivial problem.

The claim not only has a weak theoretical foundation, it fails to be supported by sufficient experimental evidence. The result of testing the algorithm on one aliased image is given; though the algorithm performs relatively well in this case, it is not clear that such behaviour can be expected in general.

One important criticism of the algorithm is that it uses only an extremely small portion of the available information. Many of the difficulties the algorithm has with noisy data could be ameliorated by sampling more of the data. Although this would increase the necessary computation, the potential benefits outweigh the negative considerations.

Krumm and Shafer present a method not dissimilar to others which have been proposed. This particular method seems unfinished; the authors themselves refer more to the potential applications and improvements permitted by the approach than to the success of the method as given.

3 Wigner distribution

Jau and Chin [4] present another method for the shape-from-texture problem. This method attempts to estimate surface orientation through local spectral analysis of image textures, with analysis begin carried out using the two-dimensional Wigner distribution. Textures are assumed to be homogeneous when viewed orthogonally to the image plane.

The Wigner distribution of a two-dimensional continuous function $f(x, y)$ is defined as

$$W_f(x, y, \theta, \phi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-j(\alpha\theta + \beta\phi)} f(x + \frac{\alpha}{2}, y + \frac{\beta}{2}) f^*(x - \frac{\alpha}{2}, y - \frac{\beta}{2}) d\alpha d\beta$$

in which x and y are position variables and θ and ϕ are spatial-frequency variables. The image energy at a location (x, y) is given by the integral of the Wigner distribution over its spatial-frequency variables, evaluated at (x, y):

$$|f(x, y)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_f(x, y, \theta, \phi) d\phi d\theta$$

while the energy at (x, y) within the bounds $\theta_a < \theta < \theta_b$, $\phi_a < \phi < \phi_b$ is given by limiting the range of integration in the previous equation:

$$E_{ab}(x, y) = \frac{1}{2\pi} \int_{\theta_a}^{\theta_b} \int_{\phi_a}^{\phi_b} W_f(x, y, \theta, \phi) d\phi d\theta$$

The image energy is used as a gauge of spatial activity at a particular location in an image. The gradient of spatial activity gives rise to surface orientation.

Since the application of the method will use discrete images with finite extent, discrete versions of the above equations are used in practice. Sufficient oversampling of $f(x, y)$ eliminates aliasing effects.

Jau and Chin observe that high-density textures have high spatial frequencies. They correlate the energy distribution to the texture gradient, thus providing a means to estimate surface orientation. A parameter used in the algorithm is the frequency threshold, a threshold determining the spatial-frequency range used to determine the energy distribution.

The equation of the surface is expressed as an equation in (x, y, z). In the implementation given, the surface is a plane; since the solution does not depend on the form of the equation, extension of the algorithm to higher-order surfaces is not difficult. An obvious extension of the work is to apply the method to nonplanar surfaces.

One approximation used is that the surface depth varies linearly with position. Provided the depth does not change much, this approximation holds. For a planar surface of limited extent this is true provided that the slant angle is not too great.

Jau and Chin test their method on several images. Their results are carefully catalogued and statistical analysis is given. They conclude that for a sufficiently large window size, the performance of the algorithm is independent of window size. They also conclude that for a frequency threshold ω in the range $0.2\pi \leq \omega \leq 1.5\pi$ the performance is independent of ω. For smaller ω the method is too insensitive to changes in spatial activity. For larger ω, the method is too sensitive to high-frequency noise.

Jau and Chin proposed to present a method for computing shape from texture; thus, the extent to which they succeed is the extent to which the method presented is applicable. Their algorithm provides good results under the restrictions they impose, although it is extremely costly to compute each estimation.

Their research methodology is sound. However, their experimental work seems unfinished; the number of different textures on which the algorithm was tested is quite small. This leaves conclusions based on their experiments somewhat questionable; Jau and Chin draw stronger conclusions than are warranted by the throughness of their work.

4 Gabor wavelets

Super and Bovik [5] present a shape-from-texture method based on extracting second-order local spectral moments.

Gabor wavelets are used to extract frequency information. The Gabor functions in two dimensions are

$$h_{uv}(x, y) = \frac{1}{2\pi\gamma^2} e^{-\frac{x^2+y^2}{2\gamma^2}} e^{2\pi i(ux+vy)}$$

These are complete, but are not orthogonal; an additional set of functions are required to compute the complete representation.

As noted by Bovik *et al.* [6], however, the objective is not to compute an exact representation but to obtain a transformation responding to regularities in pattern. Thus Bovik *et al.* suggest sampling the signal through convolution with spatially distributed Gabor wavelets. The particular distribution they advocate is such that adjacent functions on the same radial axis will intersect at half-maximum.

Most of the plane receives adequate representation in this scheme; certain regions, however, do not. Bovik *et al.* claim that this is unimportant provided that the signal is relatively noise-free, but advocate a denser sampling for noisy or natural images. Despite this suggestion, the cited sampling scheme is used by Super and Bovik in their shape-from-texture algorithm.

Bovik *et al.* provide a computational estimation procedure and a variational approach to determining frequencies from the Gabor wavelets.

The use of Gabor wavelets in extraction of local frequency information according to the method of Bovik *et al.* is designed for nonstationary smoothly coherent signals. Although applicable to a wider regime of signals, the results are most relevant for signals which have those qualities. A natural application for the method is therefore the shape-from-texture problem: such images are nonstationary unless the solution is trivial, and must be locally coherent if a solution is to be found through observation of textural variation.

In their shape-from-texture algorithm, Super and Bovik assume that the size and aspect ratio of the texture does not vary. This is a version of the homogeneity assumption.

Super and Bovik give a derivation of the way in which local image spectra back-project onto local surface spectra. This leads to a formulation in which the surface orientation at each point may be computed to be one of four possibilities. Note that the ambiguity is in the value of τ; σ is known unambiguously. The $\pm\pi$ ambiguity in tilt cannot be resolved using only textural information. The remaining two solutions differ significantly only for low slant values. Super and Bovik suggest that the solutions are identical for isotropic textures.

The algorithm computes the values $p = \frac{\partial z}{\partial x}$, $p = \frac{\partial z}{\partial y}$. Note that $\sigma = \arccos\ (\frac{1}{\sqrt{1+p^2+q^2}})$ and $\tau = \arctan\ (\frac{q}{p})$. Thus p and q offer a complete description of surface orientation aside from the ambiguities mentioned already. Super and Bovik relate the distortion of the second-order moments of the local spectra to the relative surface orientation. The method convolves the image with Gabor wavelets at specified radial frequencies. The spacing of sampling is such that the Gabor wavelets intersect at half-maximum in the frequency domain.

The assumption made to solve the resulting system of equations is that the minimum and maximum second-order surface texture moments are constant. As previously remarked, this is a variant of the homogeneity assumption and permits anisotropy.

Thus relative surface orientations may be obtained. In order to extend the algorithm to return absolute surface orientations, either the orientation at some point must be known *a priori* or further assumptions must be made. The assumption Super and Bovik suggest is that the minimum slant on the surface is zero; this assumption is valid for a wide class of surfaces, including all smooth closed surfaces. Unfortunately the assumption is violated by some simple surfaces such as planar surfaces.

Super and Bovik report that their method works for rough surfaces as well, despite the fact that one of their approximations is invalid in this context. They argue that the roughness of the surface corresponds to high-frequency noise and that the underlying low-frequency surface can still be extracted. Some examples of the algorithm's performance when applied to rough surfaces are given.

Super and Bovik base their claim on a few informal tests and on mathematical analysis. However, their statement is stronger than is warranted by the thoroughness of the research. Not only is the evidence presented purely anecdotal, it does not provide any meaningful measure of the success of the method.

Their results are presented in the form of needle diagrams, with the length of the needle proportional to the slant and the direction of the needle corresponding to the tilt. Although this presentation is simple and attractive, it defies quantitative comparison. The images analyzed are not themselves well understood, and hence the success of the surface recovery process cannot easily be measured. It does not seem that any attempt to quantify results has been made.

Despite these pessimistic words, the demonstration of Super and Bovik's method is - if not convincing - plausible. The method is one of few to attack the shape-from-texture problem for curved surfaces. The use of the Gabor wavelet is attractive due to its computational attributes and to the wide applicability of the wavelet in image analysis.

The main criticism of this approach is that it is not yet completed. The behaviour of the method should be more thoroughly investigated; claims regarding its success must be substantiated.

The result of applying the algorithm is thus a set of orientations for the surface at each image point. In theory, such a representation contains all the surface information afforded by the original image.

For a method to be shown to work, it must satisfy two conditions. First, it must be tested with a wide range of images. Second, the images on which it is tested must be well understood so that the success of the method can be quantitatively measured.

5 Comparisons of the algorithms

There do not yet exists standardized metrics for comparison of shape-from-texture algorithms. A metric could be based on the results for processing a benchmark set of images with known surface orientation; the error and time would be a measure of the performance of the algorithm. Such a metric would offer a quantitative basis for comparisons between algorithms.

The algorithms reviewed here have many qualities in common. In the underlying approach to the shape-from-texture problem they are identical, in that they approach the problem in the frequency domain rather than from any other direction.

The choice of representation for the frequency spectrum, whether Gabor wavelet, Wigner distribution, spectrogram, or some other representation, provides the first point of separation for these methods. In some cases, as in [3], the choice of representation is not critical to the success of the method. In others, quirks of the particular representation are exploited and the method is contingent on the use of that representation.

Each of the methods attacks the same problem, and they have varying degrees of success at this. Since it applies to a wide class of surfaces, making weak assumptions about the form the surface takes, the method of Super and Bovik could be said to be the most general. Inasmuch as it applies only to planar surfaces, the method of Krumm and Shafer is the least general; although Jau and Chin give a method for planar surfaces it is straightforward to extend their method to higher-order surfaces.

Super and Bovik do not present their results in a fashion conducive to quantitative comparison with other results. The angular errors reported by Krumm and Shafer and by Jau and Chin have similar magnitude; however, a naive comparison can be misleading. Krumm and Shafer use such highly restrictive assumptions that their algorithm can be expected to perform well on problems where those assumptions are valid. Conversely, Jau and Chin do not make assumptions about the form of the texture beyond its homogeneity.

Jau and Chin also provide a much more comprehensive catalogue of experimental results. Their algorithm is tested on several textures at several different angles, and the resulting combinatorics produce a large number of individual tests. Krumm and Shafer report the results of testing on a small number of images; the texture is varied, but the angles do not appear to vary much. It is worth noting that the best results of Jau and Chin are significantly better than the best results of Krumm and Shafer; this is likely due to the granularity issue mentioned earlier.

No concrete data concerning computational time are given by any of the researchers. Nonetheless, it is possible to make some remarks on this topic based on an examination of the algorithms.

The Krumm and Shafer algorithm as implemented is likely significantly faster than either of the others, especially for large images. This is due to the simple fact that only a minute fraction of the image data is used by that algorithm. This discrepancy in speed would only be increased if the suggestions concerning intelligent search were to be acted on.

Both of the other two methods are computationally intensive. Of the two, the method of Super and Bovik is likely to require more time; both methods would require comparable effort to compute the spatial-frequency information, but computation of the local spectral moments and local surface orientation is more expensive than solution of the least-squares problem by Jau and Chin's method. This may not remain true if the surface fit by the method of Jau and Chin is sufficiently complex.

Each of the three algorithms uses some form of the homogeneity assumption for surface texture. The Krumm and Shafer algorithm requires more than merely homogeneous texture, however. This algorithm requires that the texture must be completely stationary when orthogonally viewed.

The method of Super and Bovik casts its assumption in a different form than the standard homogeneity assumption. It is not clear that the minimum and maximum second-order moments are constant for all homogeneous textures, nor that all textures with those properties are homogeneous, though there is a large degree of correlation.

The method of Jau and Chin depends on the homogeneity of the texture. Due to the large quantity of data considered, and to the statistical fitting methods employed, the algorithm is very robust with respect to noisy or natural textures.

One issue which is often significant in image processing is aliasing. Krumm and Shafer give an explicit mechanism for treating aliased signals; unfortunately, this scheme does not much mitigate the problem for an automated system. An alternative suggested by Jau and Chin is to oversample the image and perform computation on the oversampled data. This approach is attractive due to its simplicity and to the fact that it can be incorporated into any shape-from-texture algorithm; however, it greatly increases the computational burden.

6 Conclusion

Each of the spatial-frequency methods examined has strengths and weaknesses. There are tradeoffs to be made between several considerations, including generality, speed, complexity and accuracy.

Krumm and Shafer chose to use the spectrogram in implementing a simple and relatively fast method which gives good results in very narrow circumstances. This is an example of tradeoffs which favor speed over generality; the main virtue of the method, its speed, is only possible due to the highly restrictive assumptions. Much computation is made unnecessary by the assumptions that the texture is noise-free and that the surface is planar.

Jau and Chin use the Wigner distribution to implement a method to recover the orientation of planes from a wide variety of textures. This robust method requires the use of a great deal of computational resources.

Super and Bovik give a method using the Gabor wavelet to recover the shape and orientation of very general surfaces. This method makes atypical assumptions both about the shape of the surface and about the form of the texture.

These methods attempt to solve the shape-from-texture problem and should be rated according to the extent to which they are able to do so. The Krumm and Shafer method is applicable only to a very small subset of the general problem. The Jau and Chin method requires certain characteristics of the surface - that it can be approximated by some analytic expression - but makes weak assumptions about the nature of the texture. The Super and Bovik method is applicable to a wide range of surfaces and also makes weak assumptions about the texture, though this method is more sensitive to noise than that of Jau and Chin.

Many outstanding issues in the shape-from-texture problem exist. The methods examined attempt to address the basic problem, simplified through certain assumptions. Among these simplifications is the assumption that the image extent of the surface to be recovered is known. That is, a segmentation process is understood to operate prior to the shape-from-texture method; in the methods given this assumption is in the form of an assumption about the original image: the entire image consists of one textured surface.

Even the basic problem of recovering the orientation of a plane remains open. Still, the methods reviewed represent significant progress in that area.

None of the methods offer a complete solution to the problem, but each of them makes some contributions. Even within the scope of spatial-frequency methods there are many and diverse approaches to the shape-from-texture problem, and spatial-frequency methods offer a solid basis for future efforts in this direction.

References

[1] Witkin, A. Recovering surface shapes and Orientations from Texture. *Artif. Intell. 17*, 1981, 17-45.

[2] Boyce, W. and DiPrima, R. *Elementary Differential Equations and Boundary Value Problems.* John Wiley and Sons, New York, 1986.

[3] Krumm, J. and Shafer, S. Shape from periodic texture using the spectrogram. In *Proceedings CVPR '92*. CVPR, 1992, pp. 284-289.

[4] Jau, J. and Chin, R. Shape from texture using the Wigner distribution. *CVGIP 52,* (1990), 248-263.

[5] Super, B and Bovik, A. Shape-from-texture by wavelet-based measurement of local spectral moments. In *Proceedings CVPR '92,* CVPR, 1992, pp. 296-301.

[6] Bovik, A., Gopal, N., Emmoth, T., and Restrepo, A. Localized measurement of emergent image frequencies by Gabor wavelets. *IEEE Trans. Info. Theory 38*, 2 (Mar. 1992), 691-711.