Correspondence

Adaptive Scale Filtering: A General Method for Obtaining Shape from Texture

J.V. Stone and S.D. Isard

Abstract—**We introduce adaptive scale filtering, a general method for deriving shape from texture under perspective projection without recourse to prior segmentation of the image into geometric texture elements (texels), and without thresholding of filtered images.**

If texels on a given surface can be identified in an image then the orientation of that surface can be obtained [11]. However, there is no general characterization of texels for arbitrary textures. Furthermore, even if the size and shape of texels on the surface is invariant with regard to position, perspective projection ensures that the size and shape of the corresponding image texels vary by orders of magnitude.

Commencing with an initial set F_0 of identical image filters, adaptive scale filtering iteratively derives a set F_N which contains a unique filter for each image position. Each element of F_N is tuned to the three-dimensional structure of the surface; that is, all image filters in F_N back-project to an identical shape and size on the surface. Thus image texels of various sizes, but associated with a single spatial scale on the surface, can be identified in different parts of the image. When combined with conventional shape from texture methods, edges derived using F_N provide accurate estimates of surface orientation. Results for planar surfaces are presented.

Index Terms—**Shape from texture, filter, adaptive, scale perspective.**

I. Introduction

The problem of shape from texture[1] necessarily involves establishing a correspondence between similar 'world' entities and their counterparts in different parts of an image. For the class of methods considered here, the problem is usually posed in terms of a texture whose distribution on the surface is in some sense regular or homogeneous. One then uses an observed departure from regularity of the distribution in the image to estimate the orientation of the surface being viewed. Under perspective projection, the image entities in question vary not only in their spatial distribution, but also in orientation and scale as a function of surface orientation and image position. Thus the problem of deciding which apparently different image entities are similar on the surface and the problem of estimating surface orientation are inextricably linked.

Earlier workers have underestimated, or ignored, the problem that scale poses to establishing a surface-image correspondence. Kanatani and Chou [8] suggest several schemes for overcoming the problem of "resolution threshold and sub-texture," but implement none, and admit to potential problems. Other workers [11] restrict analysis to synthetic images that can be segmented into geometric texels. Still others restrict analysis to surfaces with relatively low slant/focal length ratio (see [13]), thereby facilitating detection of image texels. Only in [4] has an explicit attempt to address the problem of scale been made.

The following comment from [5] illustrates the problem that scale presents for shape from texture methods:

> This texture is relatively fine-grained, so for the highest slant the edge detector misses most of the edges in the upper part of the image, resulting in an under-estimation of surface slant. (p.53)

II. Shape From Texture Methods

A. Characterizing Texture

Texture is often described in terms of "texels," which are supposed to be shapes or patterns which recur on a surface. However, some surface textures (e.g., wood grain) have no texels in this sense, and others (e.g., grass) often fail to yield identifiable image texels. Nevertheless, such textures provide measurable image quantities which can be used to estimate surface orientation.

The problem then is to characterize texture in a way that is sufficiently robust to allow analysis of images with a wide variety of textures from which surface orientation can be inferred. One approach measures energy associated with Fourier components within a small band of spatial frequencies on the surface texture, or (equivalently) the continuous-valued outputs of band-pass filters [9]. Methods that take this approach [2], [6], [16] work well if there is a least one peak in the distribution of surface texture Fourier components which is well separated from neighboring peaks (this is discussed at some length in [16]), or if orthographic projection is assumed[2] [9], [16]. Otherwise, in perspective images, image peaks associated with different spatial frequencies on the surface can mistakenly be treated as if they are derived from a single surface frequency.

A more robust characterization is given in [1] where texture is defined in terms of line length. Line length can be measured using a band-pass filter even if geometric texels cannot be identified on the surface, as in the examples of grass and wood. The lines detected need not form part of conventional texels, and might be larger or smaller in scale than any conventional texels which happen to be present. On a page of print, for example, the obvious texels are letters, but if an image filter is too large to detect the outlines of individual characters, then its output might correspond to lines of print, or even paragraphs, as in a page seen at a distance.

B. Assumptions About the Distribution of Texture

All shape from texture methods begin by making some sort of assumption about the texture of the surface. They then compute surface orientation parameter values that back-project the image texture to a surface texture that comes as close as possible to satisfying the chosen assumption. In [15] Stone distinguishes methods making two different types of assumptions:

1) Value-seeking methods, of which Witkin's method [17] is an early and influential example, assume that some measure defined

1. The term "shape from texture," as it is commonly used in the literature, is somewhat misleading. It is used, as it is here, to describe methods for estimating the orientations of planar surfaces (e.g., [1], [4], [5], [13]) and it is assumed that arbitrary shapes can be approximated by collections of planar facets. However, the substantial problem of deriving a useful shape description, e.g., "cylinder with right angle bend," from a collection of facet orientations, is not often addressed under this heading, and we are as guilty as others in this regard.

J.V. Stone is a joint member of the School of Biological Sciences, and the School of Cognitive and Computing Sciences, University of Sussex, Brighton, E. Sussex, BN1 9QY England.

S.D. Isard is with the Centre for Speech Technology Research, University of Edinburgh, Edinburgh, Scotland.

IEEECS Log Number P95061.

2. Methods that rely upon this assumption implicitly assume that the perspective image formed by any camera subtends a small visual angle, thus limiting perspective effects.

in terms of the surface texture has a *particular* value. Witkin assumes that the surface texture is *isotropic*. That is, all angles of orientation are equally represented by edges on the surface. Witkin's method generates surface orientation parameter values such that the estimated surface texture (found by back-projecting the image texture) is minimally anisotropic. More generally, value-seeking methods [3], [5], [9], [15], [17] generate surface orientation parameter values which maximize the agreement between the value associated with a measure (e.g., anisotropy) of the back-projected image texture and the assumed value (e.g., 0) of that measure for surface textures.

2) Invariance-seeking methods, exemplified in [1], assume that some measure, definable over regions of the surface, has the same value for all regions. In [1], the assumption made is that line length per unit surface area is the same in all regions of the surface. There is no assumption that lines have any particular density or orientation on the surface, only that the density is the same in all surface regions. Other examples of invariance-seeking methods are described in [1], [2], [4], [8], [11], [14], [15].

A given method might fail when applied to a given image if the imaged surface does not conform to the assumptions of that method. What concerns us here, however, is that methods can fail for surface textures that do satisfy their assumptions, because perspective projection causes surface edges with similar scales, but at different depths, to have different scales in the image.

C. The Problem of Scale

A line consists of a series of edges, and each of these edges can be associated with a zero-crossing in the second derivative of the energy in a band of spatial frequencies of the image luminance function. Thus the process of identifying an edge in the image depends not only on the variation in image gray level, but also on the scale at which the image is filtered to detect edges.

For the purpose of recovering surface orientation, it is important that texels associated with a single band of *surface* spatial frequencies can be identified in different parts of the image. It doesn't matter which band of surface frequencies is chosen, provided texture at those frequencies exists. But a method which confines itself to a fixed band of *image* frequencies identifies image texture associated with surface frequencies which depend systematically on image position.

Even if we restrict our definition of texel to edges, the problem of identifying image edges that correspond to a single small band of spatial frequencies on the surface has to be addressed. To return briefly to the printed page analogy given above, it doesn't matter whether edges identified in the image correspond to text lines, paragraphs or even individual characters, provided all of the image lines are associated with only one type of surface entity. But for large values of slant the image of a character on a near part of the page might have the same width as a line of print on a more distant part of the page. Therefore, conventional shape from texture methods [1], [5], [8], [11], [17] will tend to be inaccurate for surfaces with large values of slant, where the variation (due to projection) of image texel size and orientation across the image is large.

III. Adaptive Scale Filtering

Adaptive scale filtering[3] (ASF) is a method for computing a set F_N of filters (one for each image position) such that the back-projections of all image filters are identical. We refer to such a set as an *ideal filter set*. Applying each filter in the ideal set at its corresponding point is equivalent to convolving the imaged surface with a single, fixed-sized filter. Filter output in all parts of the image therefore corresponds to the same band of surface spatial frequencies.

In describing ASF, when we refer to filtering an image with a set of filters, each filter in the set is associated with a single image position, and each filter is applied only at its associated image position.

The process of ASF is illustrated in Fig. 1. Initially each identical circular (difference of Gaussian) filter in F_0 is applied at its corresponding point in the image. Image edges are obtained from the unthresholded zero-crossings in the filtered image. These edges are then used to provide an estimate of the surface orientation T. This estimate can be obtained by any one of several shape from texture methods (e.g., [8], [1], [13]). The initial estimate T_1 is usually inaccurate because shape from texture methods rely on the assumption that texels detected in a given image are derived from a similar scale on the surface. Using the identical filters of F_0, this assumption is only true for a fronto-parallel plane.

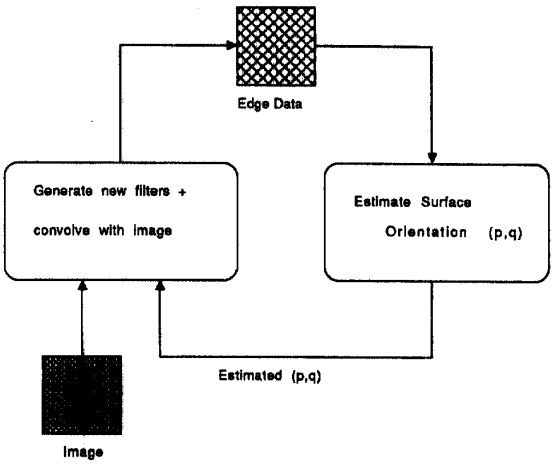

Fig. 1. Iterative process of obtaining edge data and estimating surface orientation. The filters in the first filter set are identical to each other.

Using T_1, a new set of filters F_1 is constructed. The set F_1 consists of filters, one for each pixel in the image, such that the back-projections of all filters are identical on the estimated surface with orientation T_1. Next, the image is filtered with the filter set F_1. That is, at each image pixel we apply the corresponding filter from F_1. The resultant set of edges is used to generate a new estimate T_2 of T, and then a new set of filters F_2 is computed. This procedure is repeated, and has been found to converge for all images tested so far [13]. The result is a set of image filters F_N that detects *all and only* events from a single spatial scale on the surface.

The ASF method has points in common with the technique presented by Blostein and Ahuja in [4]. They search for circular regions with uniform gray-levels at several different scales in the image. Adjacent circular regions are combined into candidate texels, and an orientation is sought which provides the best coverage of the entire image by texels of similar surface area. As with ASF, similar sized surface features project to different sized image features under perspective projection, and are detected by the use of different sized image filters.

3. In [12], [13], [14], [15] this method was referred to as adaptive *multi*scale filtering. We have changed the name to avoid confusion with methods which simultaneously filter each image point with multiple filters.

Several important differences are:

1) ASF is not a shape from texture method in itself; it is a meta-method that can be used in combination with a variety of methods which use measures of texture to estimate surface orientation. In [13] ASF has been combined with the methods described in [1], [8], [17], [5], [14], [13], which embody a number of different assumptions about surface edge distributions. For example, [1], [8] assume that edge density is invariant, whereas [5], [15], [17] make assumptions about edge orientation. The work reported here has only combined ASF with methods that use edge data to estimate surface orientation, but it could equally well be combined with methods that use other measures of surface texture (e.g., [9]). In contrast, the assumption that texel area is invariant is an integral part of the method presented in [4].
2) Blostein and Ahuja search for *closed-contour* gray-level texels, whereas we work with edges only.
3) Blostein and Ahuja employ a fixed set of image scales whereas ASF uses image scales which are iteratively adapted to suit the image being analyzed.
4) Our use of different sized image filters explicitly tracks one surface scale on all parts of the image. In [4] there is no explicit assumption that similar features in near and far parts of the scene are detected by large and small filters, respectively, although this is presumably what often happens in practice. In [4], surface orientation is determined on the basis of texel area alone, without regard to the scales of the circular regions which constitute each texel.

Adaptive scale filtering has thus far been implemented for texels that are defined as edges on a planar surface. However, the method could be implemented for other types of texel (e.g., lines, corners, geometric texels), for continuous-valued output of filters [9], [16], and for non-planar surfaces.

The following analysis assumes a planar textured surface.

A. Calculating the Dimensions of Image Filters

We require that set of image filters which would be obtained by projecting a set of circular filters from the estimated surface into the image. A small circular filter of diameter S on a surface projects to an elliptical image filter. Rather than filtering an image with elliptical image filters (as in an earlier version of ASF [12]), we model each image filter as circular. Thus we do not attempt to model the shape of image filters, but only their relative areas. This provides considerable savings in generating each filter, and we have found that the performance of the method is not noticeably altered by the use of circular filters.

In order to describe how to calculate the dimensions of image filters we first need to establish a coordinate system. The equation of a plane can be written in terms of the 'world' coordinates (X, Y, Z) as $AX + BY + CZ - D = 0$. For an imaging system with focal length, $f = 1$, and with the image plane at $Z = 1$ this can be written in terms of image coordinates $(x, y) = (X/Z, Y/Z)$ as:

$$Z = K/(1 + px + qy) \tag{1}$$

Where $K = D/C$ determines the distance along the Z axis from the plane to the origin, and $p = A/C = -\partial Z/\partial X$, and $q = B/C = -\partial Z/\partial Y$.

Our objective is to compute the radius s of each circular image filter, such that each image filter back-projects to the same amount of surface area. This can be achieved as follows. Consider the surface area A corresponding to the image area $a = \pi s^2$ of a circular image region w which is centered at a point x, y. The mapping of area from surface to image is described in terms of a *point density function pdf*:

$$A = \int_w pdf(x, y, p, q, K)\ da \tag{2}$$

where *pdf* is defined [13] as:

$$pdf(x, y, p, q, K) = K^2\left(1 + p^2 + q^2\right)^{1/2}(1 + px + qy)^{-3} \tag{3}$$

For a small circular image region centered at (x, y) we can approximate A as:

$$A \approx a \times pdf(x, y, p, q, K) \tag{4}$$

Substituting $S = (A/\pi)^{1/2}$ and $a = \pi s^2$ into (4), rearranging, and taking square roots of both sides of (4):

$$s(x, y) \approx S/pdf(x, y, p, q, K)^{1/2} \tag{5}$$

B. Simulating ASF Using Inverse Perspective

The method of ASF requires that we compute the radius of a circular filter for each image pixel. Having done so, we then need to apply each filter at its corresponding image pixel. The substantial savings that are normally made with a uniform set of circular difference of Gaussian filters (by using four one-dimensional Gaussian filters) are lost, because these savings depend upon all of the circular filters being the same.

An alternative to filtering the image with a set of filters of varying sizes is to transform the image using an *inverse perspective* transformation, and then to convolve the result with a single circular image filter. The rationale behind this method is as follows. Using ASF, each new estimate of surface orientation defines a new set of image filters. *A critical property of this set is that each filter maps to the same amount of area on the estimated surface.* However, it would be preferable if image filters back-projected so as to form surface filters with identical areas *and* shapes.

Given an estimate of the surface orientation parameters, an inverse-perspective transformation can be applied to the image to generate a fronto-parallel view of the estimated surface. This transformation consists of algorithmically rotating the estimated surface to a fronto-parallel orientation, and then reprojecting the 'rotated surface' into the image. Identical circular filters in this transformed image back-project to a set of identical circular filters on the estimated fronto-parallel surface. As with the ASF method, the accuracy of this process depends upon the accuracy of successive estimates of the surface orientation.

The value of S can be used to define a scale on the surface. A set of image filters whose sizes are a function of (5) can be made to detect surface events at the scale defined by the value of S. This method is almost identical to the camera rotation method for synthetic data described in [8], although the reason for using it is quite different.

The above process is approximately equivalent, in terms of the sets of filters applied to a textured surface, to ASF. In practice it differs from ASF in two respects: i) algorithmic rotation of a discrete image requires gray level interpolation, and this degrades the data (especially in the foreground) available to the circular filters of the inverse-perspective method, and ii) whereas the circular image filters of ASF are an approximation to the projection of circular surface filters, the process of inverse-perspective provides an image for which a set of circular image filters projects exactly to a set of circular surface filters (for a given estimate of surface orientation). In our experience, the disadvantage associated with i) appears to outweigh the advantage associated with ii), and the results from this method are inferior to those from ASF.

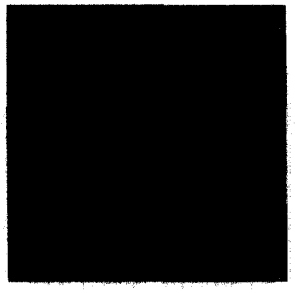

Fig. 2. Synthetic image of textured surface.

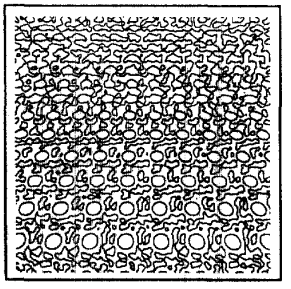

Fig. 3. Edge map from initial filtering of Fig. 2.

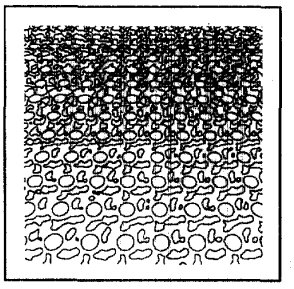

Fig. 4. Edge map from final filtering of Fig. 2.

Fig. 5. Algorithmically rotated image of boxfile.

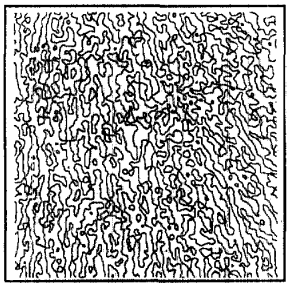

Fig. 6. Edge map from initial filtering of Fig. 5.

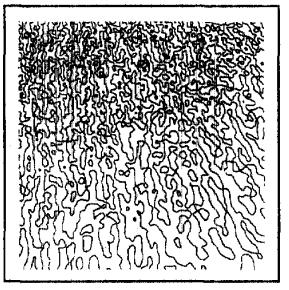

Fig. 7. Edge map from final filtering of Fig. 5.

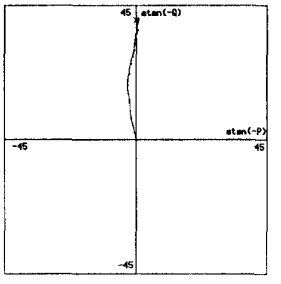

Fig. 8. Successive estimates of (P,Q) for Fig. 2.

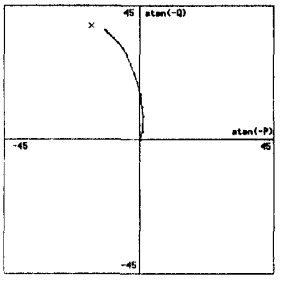

Fig. 9. Successive estimates of (P,Q) for Fig. 5.

IV. Results

Adaptive scale filtering has been tested on images of planar surfaces. Each 512 × 512 image has 256 gray-levels. The focal length of the imaging system is 512 pixel units. Each image was filtered with a difference of Gaussian (DOG) with ratio of large to small Gaussian set to 1.6. The standard deviation of the larger of the two Gaussians used to construct the DOG filter at the image center was 5 pixel units in all ASF iterations. All zero-crossings in the filtered image were labeled as edges.

ASF has been tested with modified forms of the methods described in [8](K&C), [1](ALO), and the method in [14](ABM). The modified versions of the methods in [8], [1] are described on pages 71-72 and 85-87, of the technical report cited with [13]. For the purposes of this paper, these modified versions are basically the same as those described in [8], [1]. Other shape from texture methods (see [13]) have also been tested using other images, with similar results to those presented here. Detailed results are presented for method K&C. This method assumes that the texture on the surface is homogeneous in a sense which amounts informally to assuming that the distribution of surface *texture* in the image is the same as the distribution of surface *area* in the image.[4] As described in [8], an iterative search is performed for surface orientation parameter values which project the image texture to a surface texture having this property.

The filter size at the center of the image remains constant across all iterations; this is achieved by setting $K^2(1+p^2+q^2)^{1/2}=1$. The size of image filters varies with respect to this filter according to the estimated surface orientation.

Figs. 2 and 5 depict images of textured planes, with orientations $T=(p, q)=(0.0, -0.839)$ and $(0.287, -0.788)$, respectively. Fig. 2 is synthetic, and Fig. 5 was formed by algorithmically rotating the image of a fronto-parallel textured surface. This image of the side of a boxfile was obtained using a CCD camera.

Tables I and II list the p and q values produced by ASF for Figs. 2 and 5, respectively, in conjunction with the three methods K&C, ABM and ALO, described above. The error is defined as the angle (in degrees) between the estimated and actual surface normal of a given surface.

TABLE I
Results for Fig. 2 Using Different Shape From Texture Methods

Method	p	q	Error
K&C	0.010	-0.863	0.87°
ABM	-0.047	-0.850	2.09°
ALO	-0.14	-1.08	7.17°

The error is the angle (in degrees) between the estimated and actual surface normal.

TABLE II
Results for Fig. 5 Using Different Shape From Texture Methods

Method	p	q	Error
K&C	0.209	-0.756	3.43°
ABM	0.090	-0.837	8.89°
ALO	0.105	-0.301	28.22°

The following refers to results obtained using method K&C([8]), but the general pattern of results described here is similar for other methods (see [14], [13]). Figs. 3 and 6 show the result of filtering with a set F_0 of identical filters. This is equivalent to convolving the

4. The method assumes that the first moment vector of *surface texture* around the image origin is the same as the first moment vector of the *surface area* around the image origin.

image with a single filter, and represents the data available to a standard shape from texture method not using ASF. Image edges reflect activity across a range of spatial scales on the respective surfaces. The estimates of surface orientation T_1 based on these edge maps, are consequently in considerable error. The difference between the actual and estimated surface normals is 32.72° and 37.72°, respectively, (T_1 = (0.012, −0.128) and (0.014, −0.048) for Figs. 2 and 5, respectively). Since these results are based on image edge data obtained from a conventional image filtering operation, they represent the estimates of surface orientation of method K&C using unthresholded edge data, and without ASF.

For each image, a new set of image filters F_1, based on the value of T_1, is constructed and applied to the original image. This new set of image filters is constructed according to the method outlined in the previous section, yielding a new filter for each image position. This process of filtering and re-estimation of T is repeated and converges at T = (0.01, −0.863) (Fig. 2), and (0.209, −0.756) (Fig. 5). These represent errors in the surface normal of 0.87° and 3.43°, respectively. The edge maps from the final filtering operations are shown in Figs. 4 and 7. Plots of estimated orientation versus iteration for both Figs. 2 and 5 are given in Figs. 8 and 9, respectively. The estimated surface orientation at each iteration is marked on the curve of each plot, and the actual surface orientation is marked with a cross. In both cases, using edge data generated by filtering the images with the filter set associated with the actual orientation (the ideal filter set) produces a negligible increase in accuracy, suggesting that the final edge maps are as good as they can be using a difference of Gaussian filter.

V. Discussion

For Figs. 2 and 5, edges associated with many spatial scales on the surface were initially detected. Conventionally, a proportion of such edges would be discarded on the grounds that they could not be incorporated into image texels [11] or lines [1], [5]. In contrast, each edge in all edge maps of both images tested here was given equal weighting in estimating T. Thus even the ill-defined surface objects corresponding to the image edges of Fig. 3 contribute to the estimation of surface orientation. Note that ASF works well on an image (Fig. 5) for which no patterned texels can be identified.

A conspicuous feature of the final edge maps is that they do not necessarily display perceptually salient features of the image data. Thus the final set of filters used to analyze Fig. 2 does not detect only the circular texels apparent in the image. This is because the ASF method works, not by detecting salient surface features, but by detecting any features which are reliably associated with a single scale on the surface. The method does not clean up, or segment, the image; it only finds that filter set which, for a given image, provides a stable estimate of surface orientation from one iteration to the next. The filtering operations are anchored to a particular scale S, on the surface. If edges associated with that scale can be detected in all surface regions, and the distribution of the edges satisfies the assumptions of the shape from texture method being used, then the surface orientation can be obtained using ASF. This is true even for textures which contain almost no perceptually salient features, as in Fig. 5.

As we remarked earlier, ASF can be used in conjunction with many different shape from texture methods. Here, ASF was paired with several such methods [1], [8], [14], and in [13], [14] ASF has been applied to a variety of images, synthetic, algorithmically rotated, and natural. In all cases it converged, although not necessarily to a correct answer if the surface texture violated assumptions associated with the shape from texture method used. It is difficult to say anything very general about convergence because it appears to depend on the interaction between ASF itself, the shape from texture method used, and the different ways in which particular images can satisfy, or fail to satisfy, the assumptions of particular methods.

A. Choosing a Surface Anchor Scale

In both test images (and in all images we have tested) the size of the surface scale is anchored to a fixed filter size at the image origin. Choosing a fairly arbitrary fixed value of S provides a test of the ability of ASF to converge on a set of surface features associated with that scale, even though S may not coincide with a peak in the Fourier transform of the surface texture.

A possible improvement would be to set the surface scale parameter S so that the characteristic frequencies of the final image filter set correspond to a high energy surface Fourier component. This could be achieved by using data from the first set F_0 of identical image filters to find that image region w_h associated with the highest absolute filter response. The surface orientation T_1 estimated from the data provided by F_0 can then be used to estimate, via (5), the surface scale S_h associated with the high energy Fourier component detected in w_h. Now, instead of using the original value of S to define the anchor filter size in the image, we could generate the second set of filters according to:

$$s(x,y) \approx S_h \Big/ pdf(x,y,p_1,q_1,K)^{1/2} \tag{6}$$

This would ensure that each image filter in F_1 back-projected to the same size as the filter of w_h in F_0 (thus the filter of w_h remains unchanged). This, in turn, means that each filter in F_1 back-projects to a surface scale which is closer to S_h than the corresponding filter in F_0. The procedure described for adapting the scale factor S_h between F_0 and F_1 can be repeated for subsequent iterations.[5]

B. Parallel Multiscale Filtering

The question that originally motivated this work was: How might an area of visual cortex, with its multiplicity of receptive field types, compute the orientation of a textured surface? In terms of Marr's [10] three levels of analysis (computational, algorithmic, implementational), the method of ASF is pitched at the algorithmic level; it specifies a method for executing a particular computational task. The task consists of deriving a set of image filters F, appropriate to a particular surface orientation T, where the values of F and T are initially unknown. There are many algorithmic level descriptions for executing this task, of which is ASF is only one.

In order to estimate surface orientation T, a distribution of image filters F which can identify events derived from a small band of spatial frequencies on the surface must be established. F and T are co-determined, and in ASF improved approximations of each are used to obtain improved approximations of the other. However, the fact that F and T are co-determined does not mean that they cannot be evaluated independently (e.g., T could be obtained from stereo information). In fact, we now argue that it is possible to evaluate F independently of T, and then to use F to obtain the value of T.

Consider a one-dimensional (1D) image of a 1D textured surface. In this case we require only one parameter, say q, to specify the surface orientation. The ratio of image length to surface length is proportional to $((1 + qy)/K)^2$. Accordingly, in Fig. 10 the extent of image texels increases from left to right.

5. In subsequent iterations it would not be correct to use the image region w_h with the highest filter response because an image luminance correction would have to be made, according to the current estimate of the plane's orientation.

6. This is essentially the arrangement used in [4], and adopting it would eliminate the third of the four differences listed in section 3 above.

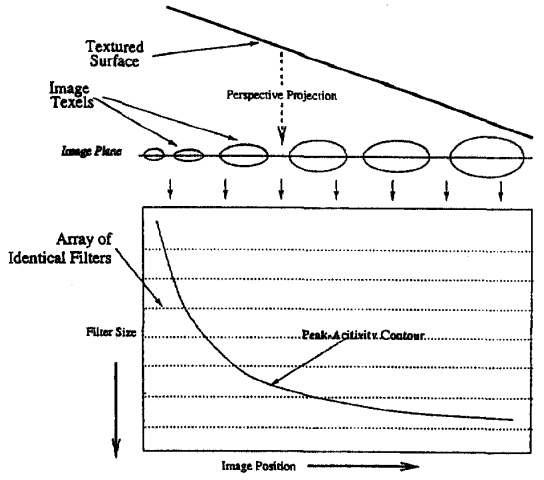

Fig. 10. A 1D textured surface projected onto an image plane produces peak-activity contours in stacked arrays of filters.

The required distribution of image filters F can be obtained from a bank B of stacked arrays of filters, where each array consists of a set of filters of a single size, and each image point is analyzed at many scales (i.e., by each level in the bank of filters).[6]

Let a filter at image position y, and with size parameter σ, be $f(y, \sigma)$. For each image position y there is a corresponding filter $f(y, \sigma)$ which represents a peak or local maximum of activity in the filter bank B. If we trace a curve through the set of peaks in the filter bank (where each peak corresponds to a local maximum of filter activity) it would look something like the curve drawn in Fig. 10. Moreover, the distribution of filters specified by such a curve is *exactly* the distribution F of filters sought by the ASF method. If there are many peaks in the Fourier transform of the surface texture then there will be many parallel *peak-activity contours* in the filter bank. In summary, each peak in the Fourier transform of the surface texture gives rise to a peak-activity contour in B; and each of these yields a filter set F, any one of which may be used to estimate the surface orientation T.

The method of ASF was born partly from the observation that a peak-activity contour in a bank of filters specifies the orientation of a textured surface. In a similar vein, Johnston [7] has independently proposed that the known distribution of cortical receptive field sizes and densities may facilitate visual interpretation of textured surfaces:

> Because of the resolution threshold problem, the uniformity [homogeneity] of an oriented surface can only be detected for paths [peak activity contours] along which the density of texture matches the scale of spatial analysis. (p. 11)

VI. Conclusion

Adaptive scale filtering provides a general method for deriving shape from texture without recourse to prior segmentation of the image into discrete texture elements, and without any form of thresholding of filtered images.

The problem of scale is an integral part of the problem of shape from texture. The process of adaptive scale filtering treats it as such, yielding accurate estimates of surface orientation even for images of surfaces with ill-defined textures.

Acknowledgment

Thanks to Raymond Lister for comments on drafts of this paper, and to the referees for their comments.

References

[1] J. Aloimonos, "Shape from texture," *Biological Cybernetics*, vol. 58, pp. 345-360, 1988.

[2] R. Bajcsy and L. Lieberman, "Texture gradients as a depth cue," *Computer Graphics and Image Processing*, pp. 52-67, 1976.

[3] A. Blake and C. Marinos, "Shape from texture: Estimation, isotropy and moments," *Artificial Intelligence*, vol. 45, pp. 323-380, 1990.

[4] D. Blostein and N. Ahuja, "Shape from texture: Integrating surface element extraction and surface estimation," *PAMI*, vol. 1, no. 12, pp. 1233-1251, 1989.

[5] J. Gårding, "Shape from surface markings," PhD thesis, Department of Numerical Analysis and Computing Science, University of Stockholm, Sweden, 1991.

[6] J.Y. Jau and R.T. Chin, "Shape from texture using the Wigner distribution," *Computer Vision Graphics & Image Processing*, vol. 52, pp. 248-263, 1990.

[7] A. Johnston, "The geometry of the striate topographic map: Functional implications," *Models of Brain Function*, R.M.J. Cotterill, Ed. Cambridge: Cambridge University Press, 1989.

[8] K. Kanatani, T. Chou, "Shape from texture: General principle," *Artificial Intelligence*, vol. 38, no. 1, pp. 1-49, 1989.

[9] P. Kube, "Using frequency and orientation tuned channels to determine surface slant," *Eighth Ann. Conf. Cognitive Society*, pp. 235-244, Amherst, Mass., 1986.

[10] D. Marr, *Vision*. San Francisco: WH Freeman, 1981.

[11] Y. Ohta, K. Maenobu and T. Sakai, "Obtaining surface orientation from texels under perspective projection," *Proc. Seventh IJCAI*, Vancouver, Canada, pp. 746-751, 1981.

[12] J.V. Stone, "Shape from texture: Textural invariance and the problem of scale in perspective images of textured surfaces," *Proc. British Machine Vision Conf.*, Oxford, England, pp. 181-187, 1990.

[13] J.V. Stone, "Shape from texture: A computational analysis," PhD thesis, Experimental Psychology, University of Sussex, England, 1991 (also published as Technical Report 209, Cognitive and Computing Sciences, University of Sussex, 1991).

[14] J.V. Stone, "The adaptive bisector method: Separating slant and tilt in estimating shape from texture," *Proc. British Machine Vision Conf.*, Leeds, England, pp. 177-186, 1992.

[15] J.V. Stone, "Shape from local and global analysis of texture," *Philosophical Trans. Royal Society London*, Series B, pp. 53-65, 339:1287, Jan. 1992.

[16] M.R. Turner, G.L. Gerstein and R. Bacjsy, "Underestimation of visual texture slant by human observers: a model," *Biological Cybernetics*, vol. 65, pp. 215-226, 1991.

[17] A.P. Witkin, "Recovering surface shape and orientation from texture," *Artificial Intelligence*, vol. 17, pp. 17-45, 1981.