Shape Representations from Shading Primitives

John Haddon and David Forsyth

Computer Science Division
University of California, Berkeley CA 94720
Phone: +1-510-643-6763, +1-510-642-9582
Fax: +1-510-643-1534
haddon@eecs.berkeley.edu, daf@cs.berkeley.edu

Abstract

Diffuse interreflections mean that surface shading and shape are related in ways that are difficult to untangle; in particular, distant and invisible surfaces may affect the shading field that one sees. The effects of distant surfaces are confined to relatively low spatial frequencies in the shading field, meaning that we can expect signatures, called shading primitives, corresponding to shape properties. We demonstrate how these primitives can be used to support the construction of useful shape representations. Approaches to this include testing hypotheses of geometric primitives for consistency with the shading field, and looking for shading events that are distinctive of some shape event. We show that these approaches can be composed, leading to an attractive process of representation that is intrinsically bottom up. This representation can be extracted from images of real scenes, *and that the representation is diagnostic.*

1 Background

Changes in surface brightness are a powerful cue to the shape of a surface; the study of extracting shape information from image shading starts with [12] and is comprehensively summed up in Brooks' book [13]. The approach views shading as a local effect; surface brightness is modelled as a product of a visibility term and a non-negative function of the Gauss map, leading to a partial differential equation—the *image irradiance equation*—which expresses the relationship between surface geometry and image brightness. Shape from shading theories that view shading as a local effect are now widely agreed to be unsatisfactory, for three reasons: the local shading model omits the effects of diffuse interreflections, a source of substantial effects in the brightness of surfaces; the underlying shape representation, a dense depth map, contains excess detail for most recognition applications; and the necessary assumptions are unrealistically restrictive. New models of shape from shading can be obtained by changing either the type of shape representation sought in the shading field [9], or the model of shading [18, 19, 16].

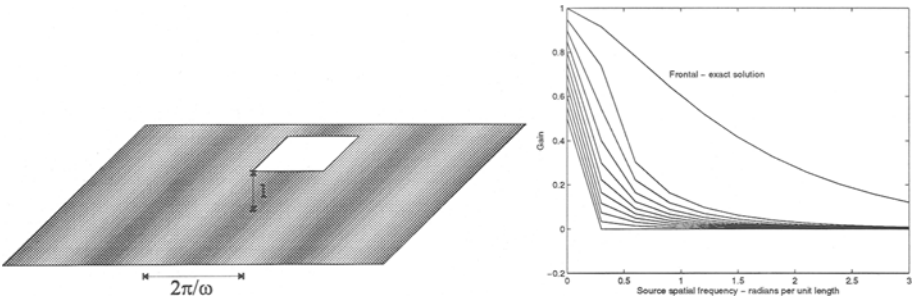


Fig. 1. *A patch with a frontal view of an infinite plane which is a unit distance away and carries a radiosity* $\sin \omega x$ *is shown on the left; this patch is small enough that its contribution to the plane's radiosity can be ignored. If the patch is slanted by* σ *with respect to the plane, it carries radiosity that is nearly periodic, with spatial frequency* $\omega \cos \sigma$*. We refer to the amplitude of the component at this frequency as the gain of the patch. The graph shows numerical estimates of the gain for patches at ten equal steps in slant angle, from 0 to* $\pi/2$*, as a function of spatial frequency* on the plane*. The gain falls extremely fast, meaning that large terms at high spatial frequencies must be regional effects, rather than the result of distant radiators. This is why it is hard to determine the pattern in a stained glass window by looking at the floor at foot of the window.*

1.1 Distant surfaces and their effects

Very few techniques for extracting shape information from shading fields are robust to the effects of diffuse interreflections—some approaches appear in [28, 21, 27, 29]. A problem arises outside controlled environments, however, because there may be surfaces that are not visible, but radiate to the objects in view (so called "distant surfaces"). Mutual illumination has a characteristic smoothing effect; as figure 1 shows, shading effects that have a high spatial frequency and a high amplitude generally cannot come from distant surfaces.

The extremely fast fall-off in amplitude with spatial frequency of terms due to distant surfaces (shown in figure 1) means that, if one observes a high amplitude term at a high spatial frequency, *it is very unlikely to have resulted from the effects of distant, passive radiators* (because these effects die away quickly). This effect suggests that the widely established convention (e.g. [3, 14, 17]) of classifying effects in shading as due to reflectance if the spatial frequency is high ("edges") and the dynamic range is relatively low, and due to illumination otherwise, can be expanded. There is a mid range of spatial frequencies that are largely unaffected by mutual illumination from distant surfaces, because the gain is small. Spatial frequencies in this range cannot be "transmitted" by distant passive radiators unless these radiators have improbably high radiosity. As a result, spatial frequencies in this range can be thought of as *regional properties*, which can result only from interreflection effects within a region.

2 Primitives

Object representation is a fundamental problem in recognition tasks. In particular, one would like to have some ability to abstract objects—recognise them at a level above that of specific instances. The classical approach to alleviating difficulties with abstraction is to view recognition in terms of assemblies of stylised primitives. In this view, which has been espoused in a variety of forms [1, 2, 20, 22], objects are represented as assemblies of shapes taken from a collection of parametric families with "good" properties. A classical difficulty with this view of representation is that it is hard to know what the primitives should be.

One important feature of geometric primitives is that their **appearance is stereotyped**. In particular, the most useful form of primitive is one where it is possible to test an assembly of image features and say whether it is likely to have come from a primitive or not. A second feature of a useful primitive is that it is **significant**. For example, a cylinder is a significant property, because many objects are made of crude cylinders. A third useful property is **robustness**; cylindrical primitives are quite easy to find even in the presence of some deformations.

In the work described in [8], it was shown that viewing objects as assemblies of primitives can be used successfully, if crudely, to find images containing horses. The program first finds the primitives—in this case, cylindrical body segments, which appear in an image as regions that are hide-like in colour and texture and have nearly parallel and nearly straight sides—and then tries to form assemblies of the primitives that are consistent with the animal's joint kinematics. Our horse finder has low recall—about 15%—but marks only 0.65% of pictures without horses, and has been extensively tested on a large set of images [7].

The weakness in this program lies in the fact that there are so many sets of nearly-parallel, nearly-straight edges (potentially body segments) that, if the number is not reduced, the kinematic tests become overwhelmed. For the horse finder, this problem can be alleviated by requiring that only segments that have hide-like pixels in the interior could be body segments. This approach can be made more general, by considering the fact that shading across a cylinder-like surface is quite constrained.

2.1 Shading primitives

Traditional shape from shading requires an impractical local shading model to produce a dense depth map. For our purposes a dense depth map is heavily redundant—instead, we will concentrate on finding stylised events in the shading field that are strongly coupled to shape, which we call *shading primitives*. In [16], Koenderink observed that deep holes and grooves in surfaces have characteristic shading properties—they are usually dark, because it is "hard" to get light into them. This is clearly an important component of the appearance of surfaces. For example, the lines on human foreheads—geometrically so trivial that they tend not to appear in depth maps—are easily visible and used by humans for communication *because they almost always have a small attached*

shadow, which gives them high contrast. These shadows are largely independent of the details of the local shape of the surface—a deep groove will be dark, and the shape of the bottom of the groove is irrelevant. The appearance of grooves is stereotyped—grooves almost always appear in images as narrow, dark bars—and so they are easily found. This combination of significance, robustness, and stereotypical appearance is precisely what is required from a primitive.

There are two forms of test in which a shading primitive might appear. In the first case, one uses shading to test an hypothesis about shape; the test must be constructed to be robust to light reflected from distant surfaces, and to yield useful results. As we show, tests meeting these requirements can be built, *because one knows what kind of shape is expected.* In the second case, the shading is the primary object that establishes the hypothesis; for example, grooves have a characteristic appearance that can be found using a template matching like approach. Typically, complex objects will require multiple tests, and we show in section 4 that one can build composite representations using shading primitives.

3 Shading on a primitive

Cylinders are natural primitives for programs that attempt to find people or animals, like the horse-finding program above. The geometric approach to finding image regions that could represent cylinders involves finding boundaries, constructing local symmetries between boundary points (as in [4, 13]) and then constructing collections of symmetries that have the same length and whose centers lie roughly on a straight line to which they are roughly perpendicular.

In this case, exploiting shading is easy because there is already an hypothesis as to the underlying geometry (as in [10]). In particular, we can test the shading along a symmetry to see whether it is consistent with the shading across a limb.

3.1 Method

Testing whether the shading across a symmetry represents the shading across a limb cross-section requires a classifier of some form. To determine this classifier, we developed a simple geometric model of a limb cross-section, and then applied a simple shading model to the limb model to generate typical shading cross-sections. We then used these analytically determined shading cross-sections to train a classifier. Passing the segment under test to the classifier tells us whether the shading is consistent with that on a limb.

The geometrical model of a limb is approximately cylindrical, with a few variations. The cross-section of the limb is taken to be elliptical, with a randomly chosen aspect ratio, and the major axis at any angle to the observer. Since limbs are certainly not perfectly elliptical in cross-section, we add a couple of bumps or grooves to the surface. Using our shading model, we calculate the shading distribution on this shape as in figure 2. It is these theoretical predictions of shading, rather than experimental data, which are used to train the classifier. However, the theoretical model does have some parameters, such as the range

of aspect ratios, and number and size of bumps, which were tuned to give a reasonable match to the experimental data.

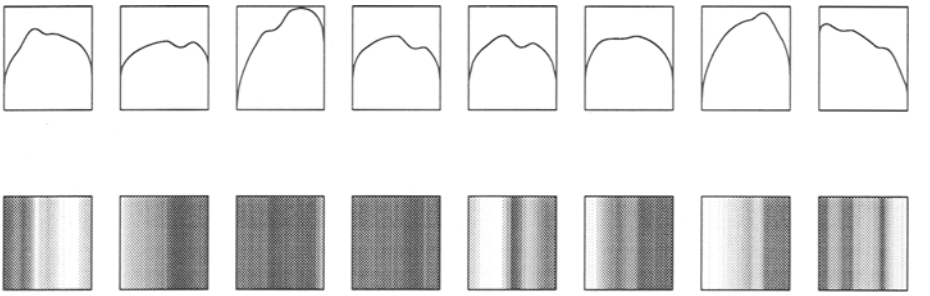


Fig. 2. *Typical limbs from our model. In each case, the plot shows the upper cross-section of the limb, while the image below it shows the shading that will be result on a limb of that shape. The bumps on the surface are intended to capture muscle definition.*

To predict the shading on our geometrical limb model, we use the same shading model as in [11]. The radiosity at any point on the limb is modelled as the sum of two components: the first due to distant radiators, which is uniform (because any spatial frequency high enough to be non-uniform over the support of the cylinder was suppressed by the low gain at high spatial frequencies); and the second due to a single point source, modeled as a source at infinity. This is a version of a model suggested by Koenderink [15] and also used by Langer *et al.*[18].

Because the limb has translational symmetry, we can model the "sky" (distant radiators) as an infinitely long half cylinder above the limb with its axis collinear with that of the limb. We can then write the brightness at a point $\mathbf{u}$ on the limb as:

$$B(\mathbf{u}) = \frac{E_a}{2}[\sin(\theta_1 - \theta_u) - \sin(\theta_0 - \theta_u)] + E_p \cos(\theta_p - \theta_u)$$

where θ_1 and θ_0 are the polar angles of the edges of the unobscured sky (measured from the zenith), θ_u is the polar angle of the the normal at $\mathbf{u}$, θ_p is the polar angle of the point light source, and E_a and E_p are the brightnesses of the ambient and point light sources. This simple model allows us to predict the radiosity given a particular limb shape.

In images, limbs appear in a variety of sizes. In order to compare limbs of different sizes, we linearly interpolate between the samples we have to create a cross section of a given width. We then project this cross section onto the most significant principal components of the positive training data, in order to generate a data point in a lower-dimensional feature space. In addition to the

principal components, we also consider the residual, a measure of the amount of variation in the signal which is not captured by the principal components. Signals similar to those yielding the principal components will be described fairly completely by the projection onto those principal components. However, signals unlike the positive data will not be described very well by the projection onto the principal components, and the difference between the original signal and its projection onto the space of principal components will be quite large. It is the energy of this difference which we call the residual.

For our classifier, we trained a support vector machine [5] using the projection onto principal components and the residual. In contrast to the use of SVMs in [24] and [23], we culled our positive training data from the results of our theoretical shading model applied to the geometrical limb model. Negative training data consisted of randomly oriented lines selected from randomly chosen images.

3.2 Results

In order to validate our geometric limb model, we compared the principal components of the images from the model, images of real limbs, and real images of things that aren't limbs. The principal components were ordered from most significant to least significant, and we then determined the matrix which transforms one set of principal components into another. The first n rows and columns of this matrix give the best map from the first n principal components in the first set to the first n principal components in the second set. The nth leading principal minor (the determinant of this $n \times n$ matrix) indicates the reduction in volume of a polytope in the first subspace when projected onto the second subspace. If the two subspaces are similar (so the $n \times n$ matrix is nearly a rotation) there will be very little reduction in volume, and the determinant will be close to one. If the subspaces are orthogonal, the polytope will collapse, and the determinant will be close to zero. Figure 3(a) shows the first thirty leading principal minors for the mappings between the three data sets. While the negatives and positives cease to describe the same subspace after only a few principal components, the theoretical and real positive data have a very strong correlation through fifteen principal components. This is a strong indication that our theoretical model is capturing the essential characteristics of shading on real limbs, because the principal components span the same space.

Since we are using this classifier as a tool to discard cross-sections which are apparently not from a limb, we require the false negative rate to be low—while it is always possible to discard a section at a higher level, once discarded at a low level, it will be very difficult to retrieve. Thus, we choose a 5% false negative rate on real cross-sections, which allows the classifier to reject 57% of negatives. While this is certainly not perfect, this does represent a significant reduction in the number of segments to be passed on for further analysis.

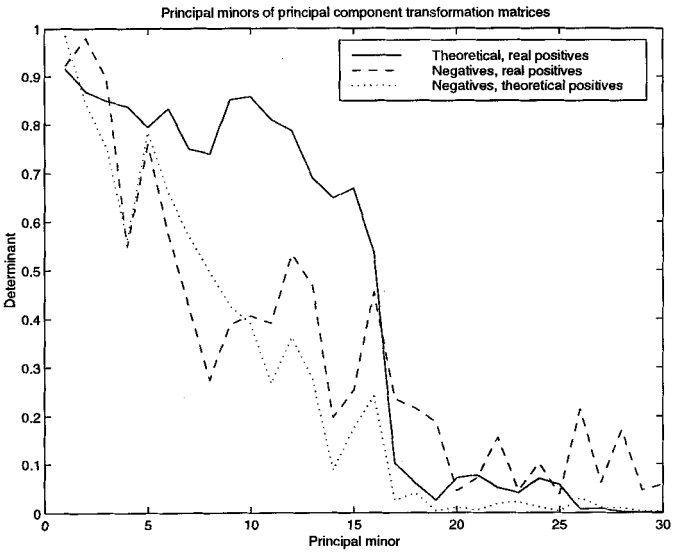


Fig. 3. *The first thirty leading principal minors of the mappings between negative, real positive and theoretical positive data. The determinant is a measure of the similarity between the subspaces described by the first n principal components in each set. The positives are very similar through the first fifteen components, while the negatives differ significantly from both positive sets after only eight components.*

3.3 Shading tests as a system component

The contribution of any visual cue should be evaluated in the context of a larger task. We have proposed to use shading cues to evaluate the hypothesis that a cylindrical primitive is present in a recognition system to find people or animals. It is natural to ask whether this improves the overall recognition process. It is difficult to give a precise answer to this question, because the learned predicates that determine whether an assembly of segments represents a person or animal are currently extremely crude. This means that we have no measure of performance that can be reliably assigned to any particular cause.

However, it is possible to estimate the extent of the contribution that testing shading makes. The standard problem with assembling symmetries is that the process produces vast numbers of symmetries, which overwhelm later grouping stages. One measure of success for measurements of shading is that they reduce this number of symmetries, without removing assemblies that could represent limbs. Since we see a shading test as more likely to be helpful in understanding large image segments (obtained using, for example, Shi's [25] normalized cut method), rather than in segmentation itself, we can apply this test on images of isolated human figures. For each of 20 images showing human figures in quite complex poses, taken from [26], we measured the rate at which the shading test rejected *symmetries* without losing body *segments*. To determine whether

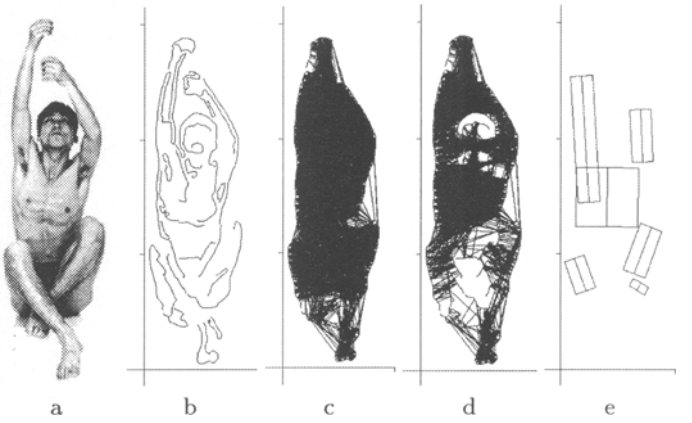


Fig. 4. *Over a test set of 20 images, symmetries are accepted by the shading test at a median rate of 61%. (a) A typical image from our test set. Notice that there is muscle definition, hair and light shadowing on the body segments, and that segments shadow other segments. The other figures on the bottom illustrate our process. Edges are shown in figure (b); figure (c) shows all symmetries found. Notice the large number of symmetries, and the spurious symmetries linking the legs. Figure (d) shows the symmetries that pass the shading test. Notice that the number of symmetries has gone down substantially, and that body segments are all represented. Figure (e) shows the segments manually determined to correspond to body segments; we have accepted that the arms, being straight, correspond to single long segments, and that one thigh is not visible as a segment, so we regard this output as containing all body segments.*

the test rejects important symmetries, we identified by hand the human body segments (upper arm/leg, lower arm/leg and torso) which did not have corresponding image segments and were visible with clear boundaries in the image. The requirement for clear boundaries ensures that errors in edge detection are not ascribed to the shading test. While this test is notably subjective, it allows some assessment of the performance of the shading cue, which is generally good—in the presence of shadows, muscle definition and the like, about half (median rejection rate is 39%) of the set of *symmetries* in a given image is rejected. The median rate at which *segments* are missed in an image is about one per two images; 10 of the images have no segments absent, five have one segment absent, and five have two absent. There appears to be some correlation with pose, which probably has to do with reduced contrast for body segments occluding other segments.

These shading tests are currently being used in a program that seeks to extract a human figure from an image. The shading test eliminates a large number of segments which are clearly not human limbs, without rejected significant numbers of actual limb segments.

4 Composite shading primitives

In [11], we developed a technique for finding grooves and folds. We applied our shading model to a geometrical model of the shape, and used these theoretical predictions to train a support vector machine to recognize grooves or folds. In that work, we were merely concerned with finding isolated shading primitives. However, difficult recognition tasks require rich representations (or, equivalently, multiple cues with multiple tests). It is therefore natural to compose tests for shading primitives. In this section, we demonstrate building a representation for a back as a near elliptical cylinder with a groove in it, by composing the tests for grooves and for limbs.

4.1 Local Properties

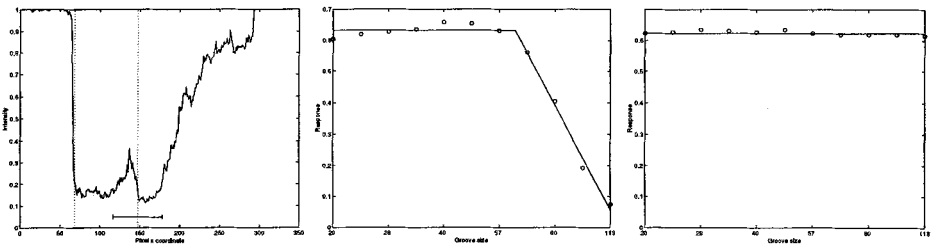
After finding the groove and localising it, we determine its width. Once we know how much of the figure has been affected by the presence of the groove, we can discount that part of the cross-section (which we do by "filling in" the groove) and can then determine whether the rest of the cross-section is consistent with the shading on a "limb".

The centre of the groove is easily found by non-maximum suppression. Currently, we only find the widths of vertical grooves, but it is easy to perform this search at arbitrary orientations, since the groove finder works at all orientations.

We search for grooves from finest to coarsest scale, linking response from scale to scale. Because the intensity pattern associated with a groove decays fairly smoothly at its boundaries, the response to a groove is essentially constant as the scale of the matching process increases, until the scale exceeds that of the groove, when the response decays slowly (see figure 5(b)). As a result, by matching from finest to coarsest scale we can reject noise responses (which do not have corresponding matches at coarser scales) and estimate the width of the groove. We fit the groove response data with two linear segments: the first, horizontal; and the second, the true line of best fit to the last values. The intersection between these lines gives us an estimate of the width of the groove.

This procedure actually improves our groove detection ability. While the groove finder does respond to the edge of a figure, it responds equally well at all scales—since there *is* no groove, it never sees the edge of the groove (see figure 5(c)). This means that, unlike real grooves, there will be no knee in the curve, allowing us to reject boundary points.

Once we have found a groove and determined its width, we can discount its effect on the shading of the back. For simplicity, we set the intensity values within the groove by linearly interpolating betwen the intensity values on either side of the groove, which gives an effect rather like filling in the groove (figure 6). While there are probably better ways of interpolating over the groove—one might use the expectation maximisation [6] algorithm to fill in this "missing" data—our approach gives perfectly acceptable results. In fact, in our current implementation, we have not actually found it necessary to discount the effect of the grooves.



(a) (b) (c)

Fig. 5. *The process of finding the width of a groove. (a) The intensity values perpendicular to the line of the groove. (b) The response of the groove detector at different scales to the groove at $x = 147$ in (a). The response is constant for small groove sizes, and then starts to drop when the size of the detector matches the size of the groove. We fit two line segments to the data, and their intersection gives the size of the groove. The calculated extent of the groove is shown by the bar in (a). (c) The response of the groove detector to the putative groove at $x = 68$ in (a). In fact, this is simply the edge of the figure, and not a groove. Thus, the groove detector has an almost constant response over all widths. Any putative groove with this signature is rejected.*

However, we expect that as our tests become more accurate, it will be necessary to account for the presence of the groove in the shading pattern across the back.

Results Figure 7 shows three typical images in the left column. In the middle column, the grey sections indicate cross-sections with limb-like shading. The top image, of a back, gives a positive response for most of the length of the back. The segments containing hair are not considered to be limb-like. The middle image is of a very flat back, with almost uniform shading, which therefore does not match the model of shading on a cylinder. It may be possible to extend our model to capture this behaviour as well. The bottom image is a fishing lure, which has many sections with limb-like shading, since its shape is roughly cylindrical, with a groove-like reflectance pattern in the centre.

4.2 Global properties

As we have seen, the shading field along a single cross-section can give us some indication as to whether the cross-section comes from a back. However, it is a much more powerful test to look at the global structure, and compare the spatial relationship of grooves and limb symmetries.

Method Because the groove detector is sensitive to orientation, we run the groove detector over the image at different orientations. Currently, we are doing this at only one scale. We find the centres of the groove by finding a high response, and stepping along the groove in the direction corresponding to the

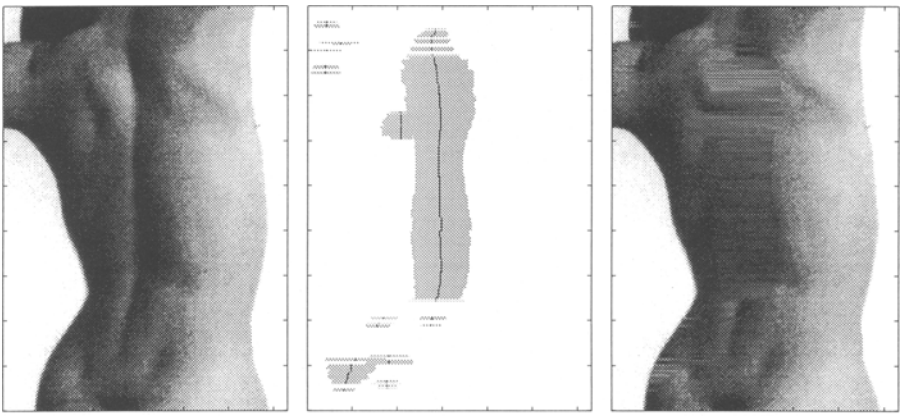


Fig. 6. *The process of groove detection and interpolation. (a) The original image. (b) Grooves in the image. The centres of the grooves are marked in black, and the widths are marked in grey. Currently, we do not follow the groove down the back while searching for widths, but it is expected that this process will allow the groove detector to jump the gaps. (c) The image with grooves filled in. The intensities at either side of the groove are interpolated linearly across the width of the groove.*

orientation with the highest response. Repeating this process until the response drops below a threshold allows us to trace out potential grooves. This process yields many potential grooves, only some of which correspond to the spine. To remove spurious grooves, we trained a support vector machine on two images, where grooves corresponding to the spine are marked as positives, and all others are marked as negatives. The features we used in the classifier were the number of points in the groove, the ratio of the number of points to the distance between the endpoints, average deviation from a straight line, and average difference between the orientation at a given point and the tangent to the groove.[1]

Using the symmetry finder discussed in subsection 3.3, we now determine which pairs of possible spine grooves and symmetry axes are consistent. The spine should be approximately parallel to the sides of the back, close to the symmetry axis, and have a region of support overlapping with the region of symmetry.

Results The spine groove classifier is effective at extracting grooves which may correspond to the spine. Out of ten images, it fails to find the spine in three cases, because the groove making up the spine is incorrectly connected to other grooves (see figure 9(a)). However, a better groove following procedure—one that tries to

[1] The orientation is determined by the maximum response at a given point, while the flow of the groove is determined by the maximum response at surrounding points.

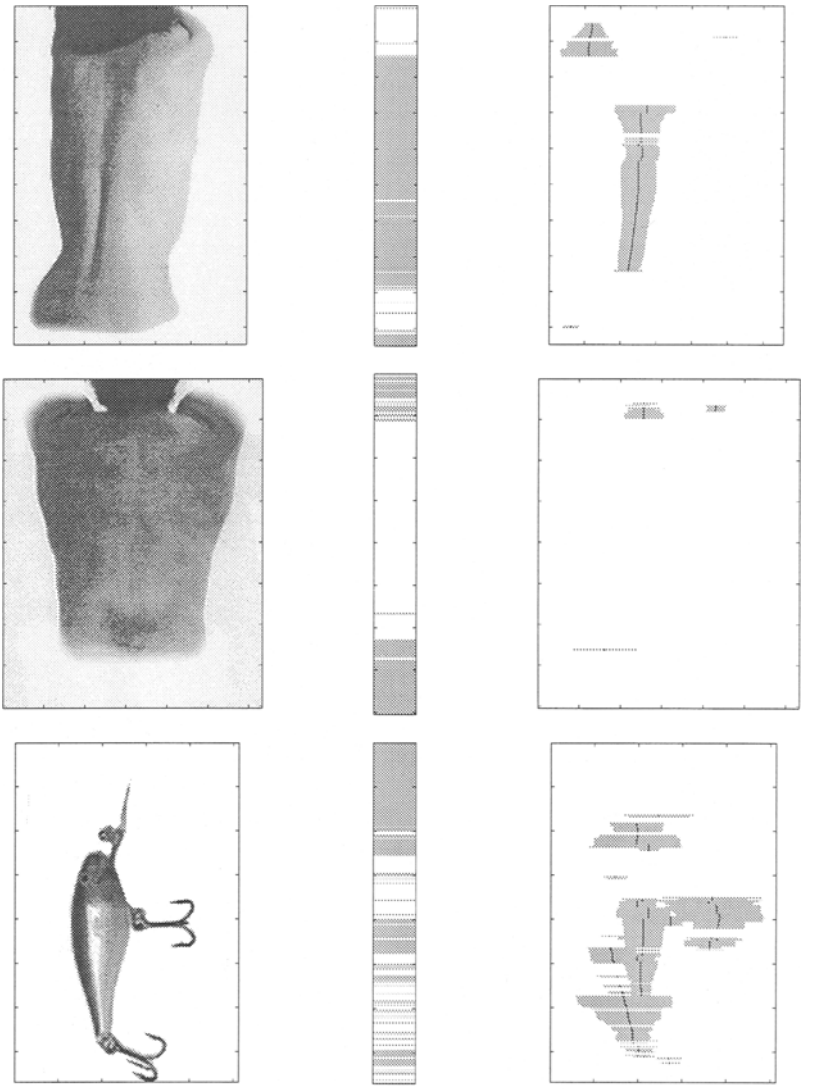


Fig. 7. *Testing cross-sections locally for shading patterns. (a) The original image. (b) Horizontal cross-sections of (a) with limb-like shading patterns are marked in grey. The breaks in the responses could be corrected by using some sort of hysteresis in the matching process. (c) Grooves in the image. The centre of each groove is marked in black, and the groove extent is marked in grey. The top row shows a back with a shading pattern consistent with the model. The middle row shows a shading pattern inconsistent with the model—the back is very flat, which creates very little variation in the intensity across the image. The bottom row shows a fishing lure, which has a shading pattern somewhat similar to many backs.*

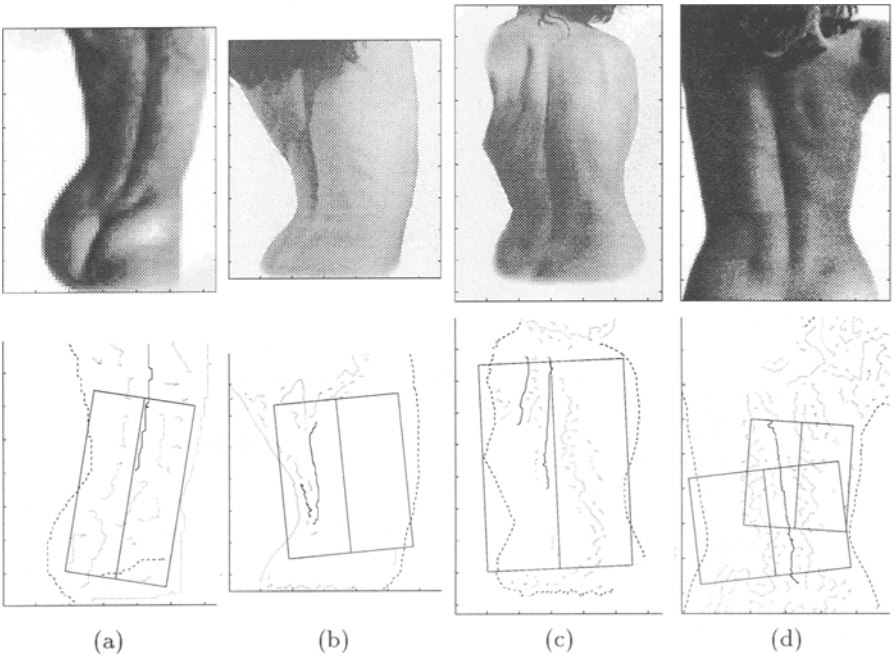


Fig. 8. *By using spatial reasoning, we can find which grooves are consistent with the groove due to the spine down the middle of the back. Top row: The original image. Bottom row: All grooves are marked with a dotted line. Grooves which could be spine grooves are marked with dashed lines. Spine grooves consistent with the axis of symmetry are solid lines. The axis of symmetry, with its length and width, are described by the rectangle.*

find straight grooves—should allow us to find the spine in these cases. In many cases, the classifier picks up the sides of the figure, since these are reasonably straight, and, out of context, are similar to spine grooves. However, these are rejected using spatial reasoning.

Overall, the conjunction of the groove primitive and the limb primitive works well. Out of seven test images in which we can find the spine, we end up with a single consistent axis of symmetry in four cases (figure 8), and two possible axes of symmetry in two more cases. In the last case, a single spurious horizontal groove allows four symmetry axes to pass the consistency tests, in addition to the two symmetry axes consistent with the actual spine groove (figure 9(d)). These horizontal symmetries, however, may be considered an artefact of the airbrushing of the image.

Out of four control images, two have one possible spine groove, which is not consistent with any axis of symmetry. In the dice example (figure 10), the edge of one die is marked as a possible groove, and is consistent with the symmetry formed by two parallel edges. However, a test that compares the orientations

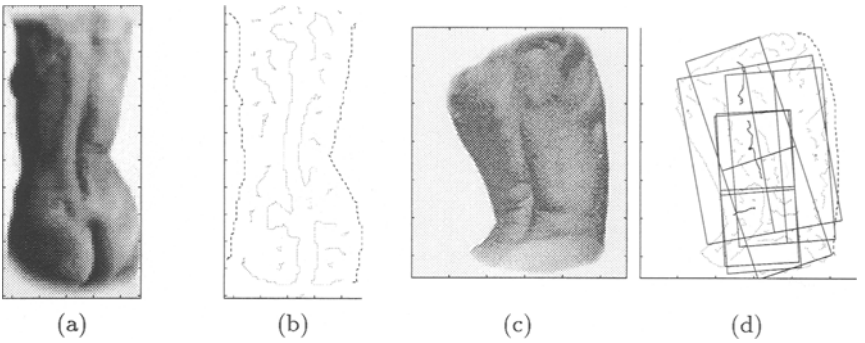


Fig. 9. *Examples of problems with the symmetry groove reasoning. Image (a) has too many other grooves, making it too difficult to find the spine groove. Image (c) has many spurious symmetries (probably due to airbrushing) and one spurious horizontal groove, causing several spurious symmetry-groove pairs. This could be rejected by examining how the shading pattern changes as one moves up and down the groove. Note that the spine grooves are marked, with the symmetry axis that would be consistent with that.*

of the grooves as compared to the orientation of the symmetry cross-sections should reject such axis-groove pairs—a true groove will have dark and light on *opposite* sides from the dark and light sides from the overall shading pattern. In this case, they are on the same side, so we should be able to reject the image.

5 Conclusions and future work

In this paper, we have demonstrated a practical use for a recognition technique based on shape from shading. Using a geometrical model of a limb, and a simple shading model, we are able to reject a large number of possible limb segments suggested by a symmetry finder. As a part of a program which finds geometric primitives and pieces them together to construct a body, this performs the valuable task of reducing the number of image segments which need to be considered as part of the kinematic chain.

Secondly, we suggested that it is possible to compose different shading primitives in order to create a more powerful decision mechanism. We showed the feasibility of composing the groove primitive with the limb primitive to get a clear description of a back.

The shading model does not take into account the effects of shadows cast by other objects. In general, it is exceedingly difficult to account for such shadows, since the object casting the shadow will not always be visible. However, the model is robust to the effects of some shadows. In figure 4, the shading test does accept even limb segments which are partially in shadow from other limb segments.

In its present form, our shading model assumes that the reflectance of the surface is approximately constant. However, the essential characteristics of shad-

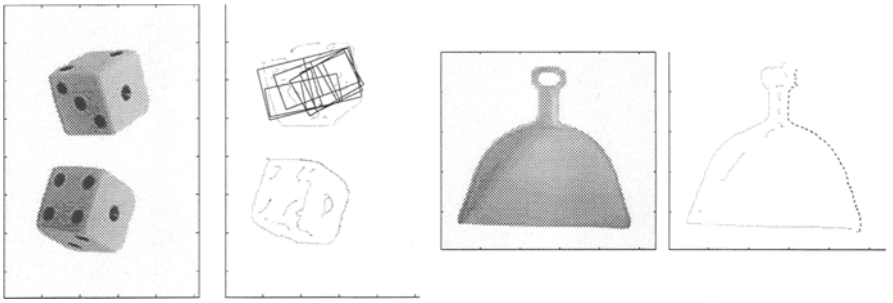


Fig. 10. *Control images. The edge of the die is marked as a possible groove, and is consistent with the symmetries found from the other edges of the die. By testing the orientations of the groove and symmetry axis, such false positives should be rejected. While the dustpan has a single groove which might be a spine, it is not in the correct position with respect to other edges in the images, and is therefore inconsistent with an image of a back.*

ing will remain across reflectance boundaries, so, in principle, there is no reason why we could not find, for example, a lycra-clad arm (since lycra is tight, the shape of the arm wearing lycra will be the same as the naked arm). Because changes in reflectance tend to be high-frequency changes, we can isolate these changes and concentrate on the mid-frequency shading effects as cues to surface shape.

Up to this point, we have demonstrated three shading primitives: folds, grooves, and limbs. We would like to extend the "shading dictionary" to include many more primitives which may be combined together to create useful, abstract representations of shape to aid in object recognition. Many shading primitives likely have very significant spatial relationships, which we would like to exploit. For example, it is relatively rare to see a single fold in clothing worn by people—typically there are several folds. (See figure 11.) Furthermore, these folds do not come in arbitrary orientations; instead, they tend to be approximately parallel. Because these spatial relationships between primitives exist, we envision a robust description of objects in terms of these groups of primitives.

The suggestion of extending the shading dictionary to include more primitives raises the question about what kinds of things are useful primitives. A useful primitive has a distinctive shading pattern which results from some class of geometric shapes. Furthermore, once one has selected a primitive, how can one best model it? The geometric models for folds, grooves and limbs all had several parameters which were tuned to give better performance. It is unclear how to tune the parameters for a given model to improve the performance.

As in all classification problems, the problem of feature selection is a difficult one. It is not clear that we have chosen the best features in this work, and the question remains as to how to select features that will best describe a given shading primitive. With the set of features we are currently using, when

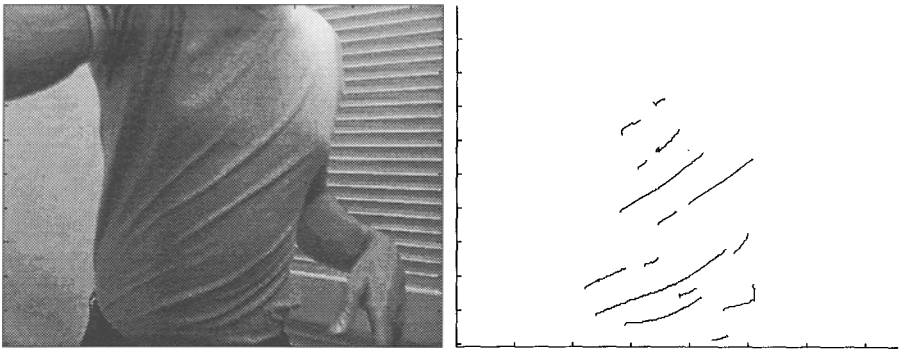


Fig. 11. *Folds in clothing have a very characteristic structure, which can be predicted from theories of buckling of shells. By grouping sets of folds with common directions, we can obtain some clue as to whether a clothed person may be found in an image. The figure on the right shows one of about twenty groups of parallel folds that are automatically extracted from the image. Note that the extent of the folds roughly corresponds to the region occupied by the torso in the image.*

the classifier gives unexpected results on given data, it can be very difficult to understand this misclassification.

References

1. I. Biederman. Recognition-by-components: A theory of human image understanding. *Psych. Review*, 94(2):115–147, 1987.
2. T. Binford. Visual perception by computer. In *Proc. IEEE Conference on Systems and Control*, 1971.
3. A. Blake. Boundary conditions for lightness computation in mondrian world. *Computer Vision, Graphics and Image Processing*, 32:314–327, 1985.
4. J. Brady and H. Asada. Smoothed local symmetries and their implementation. *International Journal of Robotics Research*, 3(3), 1984.
5. C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–97, 1995.
6. A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1977.
7. D. Forsyth and M. Fleck. Body plans. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.
8. D. Forsyth, J. Malik, M. Fleck, and J. Ponce. Primitives, perceptual organisation and object recognition. in review.
9. D. Forsyth and A. Zisserman. Reflections on shading. *IEEE T. Pattern Analysis and Machine Intelligence*, 13(7):671–679, 1991.
10. A. D. Gross and T. E. Boult. Recovery of shgcs from a single intensity view. *IEEE T. Pattern Analysis and Machine Intelligence*, 18(2):161–180, 1996.
11. J. Haddon and D. Forsyth. Shading primitives: Finding folds and shallow grooves. In *Int. Conf. on Computer Vision*, 1998.
12. B. Horn. Shape from shading : a method for obtaining the shape of a smooth opaque object from one view. Ai tr-232, MIT, 1970.
13. B. Horn and M. Brooks. *Shape from shading.* MIT Press, 1989.
14. B. K. P. Horn. Determining lightness from an image. *Computer Vision, Graphics and Image Processing*, 3:277–299, 1974.

15. J. Koenderink. Vignetting and reflexes. Shading Course Notes.
16. J. Koenderink and A. V. Doorn. Geometrical modes as a method to treat diffuse interreflections in radiometry. *J. Opt. Soc. Am.*, 73(6):843–850, 1983.
17. E. Land and J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, 1971.
18. M. Langer and S. Zucker. Shape-from-shading on a cloudy day. *Journal of the Optical Society of America A*, 11(2):467–78.
19. M. Langer and S. Zucker. Diffuse shading, visibility fields, and the geometry of ambient light. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 138–47, 1993.
20. D. Marr and K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Royal Society, London*, B-200:269–294, 1978.
21. S. Nayar, K. Ikeuchi, and T. Kanade. Shape from interreflections. *IJCV*, 6(3):173–195, August 1991.
22. R. Nevatia and T. Binford. Description and recognition of complex curved objects. *Artificial Intelligence*, 8:77–98, 1977.
23. M. Oren, C. Papageorgiou, P. Sinha, and E. Osuna. Pedestrian detection using wavelet templates. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 193–9, 1997.
24. E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 130–6, 1997.
25. J. Shi and J. Malik. Normalized cuts and image segmentation. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.
26. unknown. *Pose file 6.* unknown, unknown. A collection of photographs of human models, annotated in Japanese.
27. T. Wada, H. Ukida, and T. Matsuyama. Shape from shading with interreflections under proximal light source-3d shape reconstruction of unfolded book surface from a scanner image. In *Int. Conf. on Computer Vision*, pages 66–71, 1995.
28. R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–44, 1980.
29. J. Yang, N. Ohnishi, D. Zhang, and N. Sugie. Determining a polyhedral shape using interreflections. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 110–115, 1997.