Object Recognition Using Shape-from-Shading

Philip L. Worthington and Edwin R. Hancock

Abstract—This paper investigates whether surface topography information extracted from intensity images using a recently reported shape-from-shading (SFS) algorithm can be used for the purposes of 3D object recognition. We consider how curvature and shape-index information delivered by this algorithm can be used to recognize objects based on their surface topography. We explore two contrasting object recognition strategies. The first of these is based on a low-level attribute summary and uses histograms of curvature and orientation measurements. The second approach is based on the structural arrangement of constant shape-index maximal patches and their associated region attributes. We show that region curvedness and a string ordering of the regions according to size provides recognition accuracy of about 96 percent. By polling various recognition schemes, including a graph matching method, we show that a recognition rate of 98-99 percent is achievable.

Index Terms—Shape-from-shading, object recognition, shape-index, histograms, constant shape-index maximal patches, graph-matching.

1 Introduction

SHAPE-FROM-SHADING is concerned with recovering surface orientation from local variations in measured brightness. There is psychophysical evidence that the process plays an important role in the perception and recognition of surface topography [18], [16], [17], [1]. One of the shortcomings of the early work on shape-from-shading [7] was that it failed to provide surface information that was sufficiently accurate for the inference of surface topography or practical 3D object recognition from 2D images. The reason for this was that, by weakening the data-closeness of the image-irradiance equation in favor of smoothness, the recovery of surface detail was sacrificed.

Recently, however, there has been a consolidated effort in the literature aimed at overcoming the well-documented problems with shape-from-shading. For instance, Bichsel and Pentland [2] have a method which requires prior knowledge of the heights of singular points of the surface. Oliensis and Dupuis [19], on the other hand, have developed an algorithm which has been proven to reconstruct height information correctly from intensity data. Meanwhile, there has been interest in using the apparatus of level-set theory to solve the image irradiance equation as a boundary value problem [12], [13]. Related work by Rouy and Tourin [21] has proven the correctness and uniqueness of the solutions to shape-from-shading for the case when the light source and viewing directions are coincident. In this work, the problem is cast in a viscosity setting using the Hamilton-Jacobi-Bellman equations. In an effort to overcome the problems of poor data-closeness and oversmoothing, we have developed a new framework for shape-from-shading [26], [27], [25] in which the image irradiance equation is treated as a hard constraint [27] and new curvature consistency constraints can be used to recover meaningful topographic surface structure [26].

- *P.L. Worthington is with the Department of Computation, UMIST, Manchester M60 1QD, UK. E-mail: plw@co.umist.ac.uk.*
- *E.R. Hancock is with the Department of Computer Science, University of York, York Y01 5DD, UK. E-mail: erh@cs.york.ac.uk.*

Manuscript received 22 Mar. 2000; revised 8 Nov. 2000; accepted 17 Jan. 2001.
Recommended for acceptance by M. Shah.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 111745.

The aim in this paper is to explore whether such topographic information derived from 2D intensity images using shape-from-shading can be used for 3D object recognition. The approach has proven successful in the recognition of 3D objects from range images. For example, Dorai and Jain [4], [5] have shown how both image histograms and region segmentations, generated using the Koenderink and Van Doorn shape index [15], can be used to recognize range images. We investigate two recognition strategies. The first is a low-level method which uses histograms of image attributes derived from the needle-maps. Histograms have proven to be simple and powerful attribute summaries which can be used to great effect in the recognition of objects from large image databases. The idea was originally popularized by Swain and Ballard, who used color histograms [20], and it has subsequently been successfully used for texture recognition [6], 3D object recognition from range-images using shape-index spectra [5], and for line-pattern recognition [8]. We explore several attribute representations derived from the needle-maps returned by SFS. The attributes studied are slant and tilt angles, principal curvatures, mean and Gaussian curvature, and surface curvedness and shape-index. These are compared with intensity-based attributes.

Our second recognition strategy is posed at a higher level and uses a structural abstraction. The method follows Dorai and Jain by using "constant shape-index maximal patches" (CSMPs) extracted from the needle-maps as an abstraction of surface structure [4], [5]. We investigate several structural abstractions of the resulting topographic regions. These include the region adjacency graph and an ordered string of region attributes. The main findings reported in this paper have appeared in an abridged form in two recent conference papers [23], [24].

2 Shape-from-Shading

Our new shape-from-shading algorithm has been demonstrated to deliver needle-maps which preserve fine surface detail [27], [26]. The observation underpinning the method is that, for Lambertian reflectance from a matte surface, the image irradiance equation defines a cone of possible surface normal directions. The axis of this cone points in the light source direction and the opening angle is determined by the measured brightness. If the recovered needle-map is to satisfy the image irradiance equation as a hard constraint, then the surface normals must each fall on their respective reflectance cones. Initially, the surface normals are positioned so that their projections onto the image plane point in the direction of the image gradient. Subsequently, there is iterative adjustment of the surface normal directions so as to improve the consistency of the needle-map. In other words, each surface normal is free to rotate about its reflectance cone in such a way as to improve its consistency with its neighbors. This rotation is a two-step process. First, we apply a smoothing process to the current surface normal estimates. This may be done in a number of ways. The simplest is local averaging. More sophisticated alternatives include robust smoothing with outlier reject and smoothing with curvature or image gradient consistency constraints. This results in an off-cone direction for the surface normal. The hard data-closeness constraint of the image irradiance equation is restored by projecting the smoothed off-cone surface normal back onto the nearest position on the reflectance cone.

To be more formal, let $\mathbf{s}$ be a unit vector in the light source direction and let $E_{i,j}$ be the brightness at the image location (i,j). Further, suppose that $\mathbf{n}_{i,j}^k$ is the corresponding estimate of the surface normal at iteration k of the algorithm. The image irradiance equation is $E(i,j) = \mathbf{n}_{i,j}^k.\mathbf{s}$. As a result, the reflectance cone has opening angle $cos^{-1}E(i,j)$. After local smoothing, the off-cone surface normal is $\bar{n}_{i,j}^k$. The updated on-cone surface normal which satisfies the image irradiance equation as a hard constraint is

obtained via the rotation $\mathbf{n}_{i,j}^{k+1} = \Phi\bar{\mathbf{n}}_{i,j}^{k}$. The matrix Φ rotates the smoothed off-cone surface normal estimate by the angle difference between the apex angle of the cone and the angle subtended between the off-cone normal and the light source direction. This angle is equal to:

$$\theta = \cos^{-1} E - \cos^{-1} \frac{\bar{\mathbf{n}}_{i,j}^{k} \cdot \mathbf{s}}{||\bar{\mathbf{n}}_{i,j}^{k}|| \cdot ||\mathbf{s}||}.$$

This rotation takes place about the axis whose direction is given by the vector $(u, v, w)^T = \bar{\mathbf{n}}_{i,j}^{k} \times \mathbf{s}$. This rotation axis is perpendicular to both the light source direction and the off-cone normal. Hence, the rotation matrix is:

$$\Phi = \begin{pmatrix} c + u^2c' & -ws + uvc' & vs + uwc' \\ ws + uvc' & c + v^2c' & -us + vwc' \\ -vs + uwc' & us + vwc' & c + w^2c' \end{pmatrix},$$

where $c = \cos\theta$, $c' = 1 - c$, and $s = \sin\theta$.

The off-cone surface normal is recovered through a process of robust-smoothing. Full details of the smoothing process are outside the scope of this paper and can be found in [25].

The algorithm has a number of shortcomings. For instance, it needs known light source direction and copes neither with changes in albedo nor the presence of specularities. Much of its success can be attributed to the use of the image gradient direction to initialize the positions of the surface normals on the irradiance cone. This initialization has the effect of biasing the recovered needle-map in favor of convex surfaces. Moreover, we have no proof of the uniqueness of solutions delivered by the method. The algorithm also leaves considerable scope for further development through, for instance, the use of the integrability constraint and multiscale coarse-to-fine search. Notwithstanding these limitations, the aim in this paper is to explore the use of our shape-from-shading method for topography-based object recognition. A full analysis of the failure modes of the shape-from-shading method can be found in Worthington's doctoral thesis [28].

3 Attribute Histograms

The first aim of this paper is to evaluate the recognition performance achievable using histograms of various attributes derived from the needle-maps delivered by shape-from-shading.

The simplest attributes which can be computed from the needle-map are the local slant and tilt angles. We measure these angles relative to the mean normal direction calculated over the entire needle-map. This reference normal should ensure rotational invariance of the histogram. If the image has R rows and C columns, then the mean surface normal is $\hat{\mathbf{n}} = \frac{1}{RC}\sum_{i=1}^{R}\sum_{j=1}^{C}\mathbf{n}_{i,j}$ The slant and tilt angles relative to this reference normal are given in terms of the x, y, z components of the local surface normal and the mean surface normal by $\sigma = \arccos(\mathbf{n}_{i,j}(z)) - \arccos(\hat{\mathbf{n}}(z))$ and

$$\tau = \arctan\left(\frac{\mathbf{n}_{i,j}(y)}{\mathbf{n}_{i,j}(x)}\right) - \arctan\left(\frac{\hat{\mathbf{n}}_{i,j}(y)}{\hat{\mathbf{n}}_{i,j}(x)}\right).$$

A more complex family of curvature-based attributes can be computed from the Hessian of the surface

$$\mathcal{H} = \begin{pmatrix} \left(\frac{\partial \mathbf{n}}{\partial x}\right)_x & \left(\frac{\partial \mathbf{n}}{\partial x}\right)_y \\ \left(\frac{\partial \mathbf{n}}{\partial y}\right)_x & \left(\frac{\partial \mathbf{n}}{\partial y}\right)_y \end{pmatrix}, \quad (1)$$

where $(\cdots)_x$ and $(\cdots)_y$ denote the x and y components of the parenthesized vector, respectively. The most direct attributes available to us are the principal curvatures κ_1 and κ_2, which are the eigenvalues of the Hessian matrix and are found by solving the equation $|\mathcal{H} - \kappa\mathbf{I}| = 0$. From the principal curvatures, we can compute the mean curvature $H = \frac{1}{2}(\kappa_1 + \kappa_2)$ and the Gaussian curvature $K = \kappa_1\kappa_2$.

An alternative topographic characterisation of the surface is provided by the shape-index and curvedness representation of Koenderink and van Doorn [15]. The shape index is defined in terms of the principal curvatures using the angular measure

$$\phi = \frac{2}{\pi}\arctan\frac{\kappa_2 + \kappa_1}{\kappa_2 - \kappa_1} \qquad \kappa_2 \geq \kappa_1 \quad (2)$$

and the overall magnitude of curvature is measured by the curvedness

$$c = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}}.$$

The shape-index can be used to label the surface according to its local topography. The shape-index intervals are as follows:

- cup $\phi \in [-1, -\frac{5}{8})$,
- rut $\phi \in [-\frac{5}{8}, -\frac{3}{8})$: saddle-rut $\phi \in [-\frac{3}{8}, -\frac{1}{8})$,
- saddle-point $\phi \in [-\frac{1}{8}, \frac{1}{8})$,
- plane-undefined,
- saddle-ridge $\phi \in [\frac{1}{8}, \frac{3}{8})$,
- ridge $\phi \in [\frac{3}{8}, \frac{5}{8})$, and
- dome $\phi \in [\frac{5}{8}, 1)$.

A good review of the methods available for characterizing surface topography using techniques from differential geometry can be found in Koenderink [14].

Having defined the available attributes, we now define the histograms used in our study. We use 1D, 2D, and co-occurrence histograms. The histograms used in our study are:

1. one-dimensional gray-scale histograms,
2. one-dimensional shape-index histograms,
3. two-dimensional histograms of slant and tilt angles,
4. two-dimensional of shape-index and curvature,
5. two-dimensional histograms of maximum and minimum curvature,
6. two-dimensional histograms of mean and Gaussian curvature,
7. gray-scale co-occurrence histograms, and
8. shape-index co-occurrence histograms.

The 1D histograms each contained nine bins and the 2D histograms 9×9 bins. The co-occurrence histograms use a 5×5 support window.

We perform recognition on the basis of the minimum distance between histograms. The choice of distance metric for measuring the similarity between histograms may affect recognition performance [8]. The L1 and L2 norms are both commonly used in histogram comparison. However, Huet and Hancock [8] found that the Bhattacharyya distance offers significant advantages over the L1 and L2 norms within the context of histogram-based retrieval from large image databases.

4 Region-Based Recognition

In this section, we explore how the structural arrangement of the regions of uniform topographic class, together with region attributes derived from the needle-map, can be used for the purposes of recognition and matching.

4.1 Constant Shape Index Maximal Patches

Our region-based representation borrows some of the features of the COSMOS-representation of Dorai and Jain [4], [5], which we consider most likely to be stable when recovered using SFS.

Specifically, we use the needle-map and topographic labels in tandem to generate a rich description of object topography.

The representation describes an image of an object using a patchwork of maximal regions of constant shape-index. These constant shape-index maximal patches (CSMPs) are defined on the region of an image, O, corresponding to the object. Suppose that each pixel in the image is assigned a topographic label on the basis of its shape-index value using the intervals for the different topographic classes defined in Section 2. Let ω_p and ω_q be the topographic labels assigned to two pixels with pixel positions p and q. Further, let $\Gamma(p,q)$ be a path between these two pixels. A CSMP is a maximally-sized image patch $P \subseteq O$ such that $\forall p, \forall q \in P$, $\omega_p = \omega_q$ and there exists a connected path $\Gamma(p,q)$ from p to q consisting of points $r \in P$ such that $\omega_p = \omega_r = \omega_q$. The path condition imposes connectedness of the CSMP, defining it as a contiguous image region of constant shape index. For example, the image of a sphere should, if ideally labeled by our SFS scheme, possess a single CSMP of spherical cap shape index.

Since we are working with noisy data derived from single images, rather than the CAD data and range images investigated by Dorai and Jain [4], our regions tend to be relatively small and fragmented. To obtain a manageable list of regions, we impose a minimum region size of 25 pixels. Since the images used are of dimension 128×128 pixels, this corresponds to a limiting size of 0.15 percent of the total image area. This typically gives us between 40 and 80 regions per image on the COIL database.

Besides the CSMP region-sizes and topographic labels, we can add other attributes to the representation. For example, in the COSMOS representation, Dorai and Jain [4] incorporate the mean normal for each region. We also include the mean curvedness of each CSMP as part of the representation to give an indication of the scale of the curvature.

Potentially, the most important element of our representation is the region-adjacency graph (RAG) for the CSMPs since this encodes much of the structural information about the arrangement of the topographic structures that constitute objects in the image. Recovery of the RAG from a region-based description of an image is relatively straightforward. We opt to traverse the list of CSMPs and find all other regions possessing a thresholded number of shared border pixels. We find that a minimum of five pixels yields a detailed but manageable RAG, typically with around 10-20 adjacent regions adjoining the largest CSMPs, reducing in a well-behaved fashion to one or two adjacencies for the smallest regions, although some small regions tend to be isolated by this criterion since they do not possess sufficiently long borders with other regions.

4.2 Recognition Strategy

We adopt two different approaches to matching the CSMPs extracted from the raw shape-index delivered by the SFS scheme. The first of these is set-based and uses various attributes for the CSMPs. This first approach does not use any information concerning relational arrangement or graph-structure. The second approach is graph-based and aims to compare objects using information conveyed by the edge-structure for the region adjacency graph of the CSMPs.

4.2.1 Attribute-Based Methods

The simplest structural recognition strategy that we employ is based on the sequence-order of the normalized CSMP region areas. We define the similarity between sequences as

$$d_{cs} = \sum_{l=2}^{\min(N_M+1,N_D+1)} (A_l(M) - A_l(D))^2, \qquad (3)$$

where N_D and N_M are the numbers of regions in the data and model representations, respectively, and A_l is the normalized area of the CSMP with region label l. The label l indicates the size ordering of the region and the value $l = 1$ is reserved for the background. A more complex alternative is to characterize the regions using an attribute and to compute the similarity measure

$$d_{\alpha} = \sum_{l=2}^{\min(N_M+1,N_D+1)} \frac{(A_l(M) + A_l(D))}{2} (\alpha_l(M) - \alpha_l(D))^2, \qquad (4)$$

where α_i is the attribute associated with the CSMP with label l. In our experiments, the attribute may be 1) shape-index, 2) curvedness, and 3) mean surface normals. These measures can be viewed as a means of clustering the CSMPs. Clearly, if the region sizes are affected by segmentation error or perspective foreshortening, then we can expect the measure to fail. We attempt to limit these problems by using normalized area rather then absolute area.

4.2.2 RAG Comparison

In this section, we turn our attention to the matching of the region adjacency graph for the CSMPs. This is the most complex part of the representation and, therefore, is the most difficult and expensive part to match. Many graph-matching methodologies have been reported in the literature and it is not our intention here to investigate these. However, most of the reported methods are tailored to the problem of finding a detailed pattern of correspondences between pairs of graphs. From the standpoint of computational expense, these are not well-suited to finding the graph from a large database which is most similar to the query. Recently, Huet and Hancock [9], [8] have reported a framework for measuring the similarity of attributed relational graphs for object recognition from large structural libraries. The method uses a variant of the Hausdorff distance as a simple and efficiently computed measure of graph similarity.

The idea underpinning the Hausdorff distance, as used by Huttenlocher et al. [10] to locate objects in images, is to compute the distance between two unordered sets of observations when the correspondences between the individual elements are unknown. The method uses hard max and min operators to establish correspondences between primitives and then compute the distance between the sets. Huet and Hancock [9] have developed a fuzzy or soft variant of the Hausdorff distance which can be used as a measure of graph similarity. The measure uses a robust weighting function to compute the similarity of attributed relational graphs. The similarity measure is

$$H^r(G^d, G^m) = \sum_{(i,j)\in E^d} \left\{ \min_{(I,J)\in E^m} \rho_\sigma\left(\left\| \mathbf{v}^m_{(I,J)} - \mathbf{v}^d_{(i,j)} \right\| \right) \right\}, \qquad (5)$$

where V^d and V^m are the sets of nodes of the data and model graphs, respectively, and E^d and E^m are the corresponding sets of graph edges or arcs. The vector $\mathbf{v}^m_{(I,J)}$ is the vector of measurements associated with the graph edge $(I,J) \in E^m$ linking node $I \in V^m$ to node $J \in V^m$ in the model graph. Likewise, $\mathbf{v}^d_{(i,j)}$ is the measurement vector corresponding to graph edge $(i,j) \in E^d$ of the data-graph. For the sake of simplicity, we choose to assign the normalized region size to each node. Hence, the attribute vector of each arc in the data graph becomes $\mathbf{v}^d_{(i,j)} = \left(A_i A_j\right)^T$, where A_i and A_j are the normalized region areas represented by nodes i and j, respectively. Similar attribute vectors are defined for the model graphs, G^m.

5 Experiments

Our experimental evaluation of recognition performance is divided into two parts. We commence by considering histogram-based recognition using different attributes and different types of

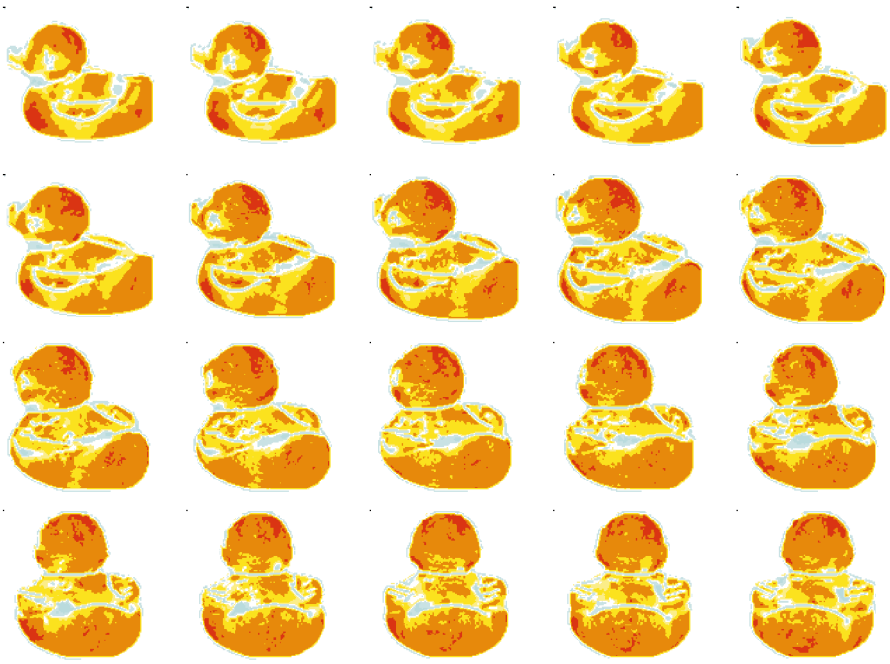


Fig. 1. As the duck rotates by 5° intervals (left to right and top to bottom), the topographic labeling remains consistent. The labels are colored according to the scheme proposed by Koenderink and van Doorn [15]. The label colors are: green = spherical cup, cyan = trough, blue = cylindrical ruts, pale blue = saddles ruts, white = symmetrical saddles, background, and planar regions, pale yellow = saddle ridge, yellow = cylindrical ridge, orange = dome, and red = spherical cap.

histogram. The second aspect of the study focuses on region-based recognition using CSMPs. The data used in our experiments is the COIL database from Columbia University. This consists of 72 views of 20 objects. The views are regularly positioned around the equator of each object, i.e., the viewing angle is incremented by 5 degrees between successive views. Each object is illuminated in the viewing direction with a parallel beam light source.

5.1 View Stability

We commence our experimental study by presenting some results which point to the stability of the histogram and region representations under varying viewpoint and illumination. To demonstrate the stability of the representation, Fig. 1 shows the CSMPs for a sequence of different views of the toy duck from the Columbia University COIL database. Notice how the valley lines around the beak and the wing are well-recovered at each viewing angle and, also, how the shape of the saddle structure below the wing is maintained. It is also illuminating to consider various region attributes which can be used to augment the representation for the purposes of recognition. Fig. 2 shows the normalized CSMP region sizes and topographic labels for the first 35 segmentations of the toy duck. The change in viewing angle between adjacent images is 5 degrees. The 25 largest regions are used in each case. Notice the stability of the areas of the different shape-index regions as the object rotates. We also illustrate the effects of variable illumination on the shape-index information returned by our new shape-from-shading scheme. Fig. 3 shows the effect of varying the light source direction on the recovered shape-index histogram. Changing the light source direction has a smaller effect on the shape-index histogram than on the gray-level histogram.

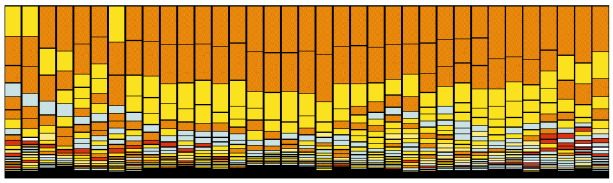


Fig. 2. A simple comparison of image structure in terms of CSMP sizes and shape-index labels. The 25 largest CSMPs of 35 images of the toy duck are sorted in order of size. Each vertical bar shows the relative sizes of these regions, colored according to their associated shape-index label. There is considerable correlation between neighboring bars in terms of both normalized region sizes and shape-index labels.

5.2 Histograms

We commence by comparing the recognition performance obtained using 1D histograms of shape-index and image intensity. The experiment is performed with the COIL database. We query the database with each image in turn. We record how many times the correct image has the closest Bhattacharyya distance to the query. The figure shows that, for the image intensity histogram, 91 percent of the queries are best-matched while, in the case of the shape-index histogram, only 73 percent are best-matched.

By contrast, if we accept the first n matches, a different picture emerges. Fig. 4 shows the recognition performance achieved by accepting different numbers of matches. If we take a sufficient number of matches, in this case, around 10, then the shape-index outperforms the gray-level histograms. In fact, it appears that, when using gray-level histograms, there are some "extreme" views of objects which possess histograms that are virtually unmatchable.

5.2.1 Alternative Attributes

We have repeated the preceding comparison for the alternative histogram representations discussed in Section 5. Fig. 5 shows the results obtained taking first n-matches. In terms of taking the single closest match, the shape-index+curvedness histogram and the histogram of the two principal curvatures perform the best. When the best n-matches are used, the 2D histograms of Gaussian and Mean curvature overtakes these to achieve 100 percent correct recognition in the first 10 matches, closely followed by the 2D histograms of shape-index and curvedness values (which performs best on first matches) and the 2D relational histograms of shape-index values.

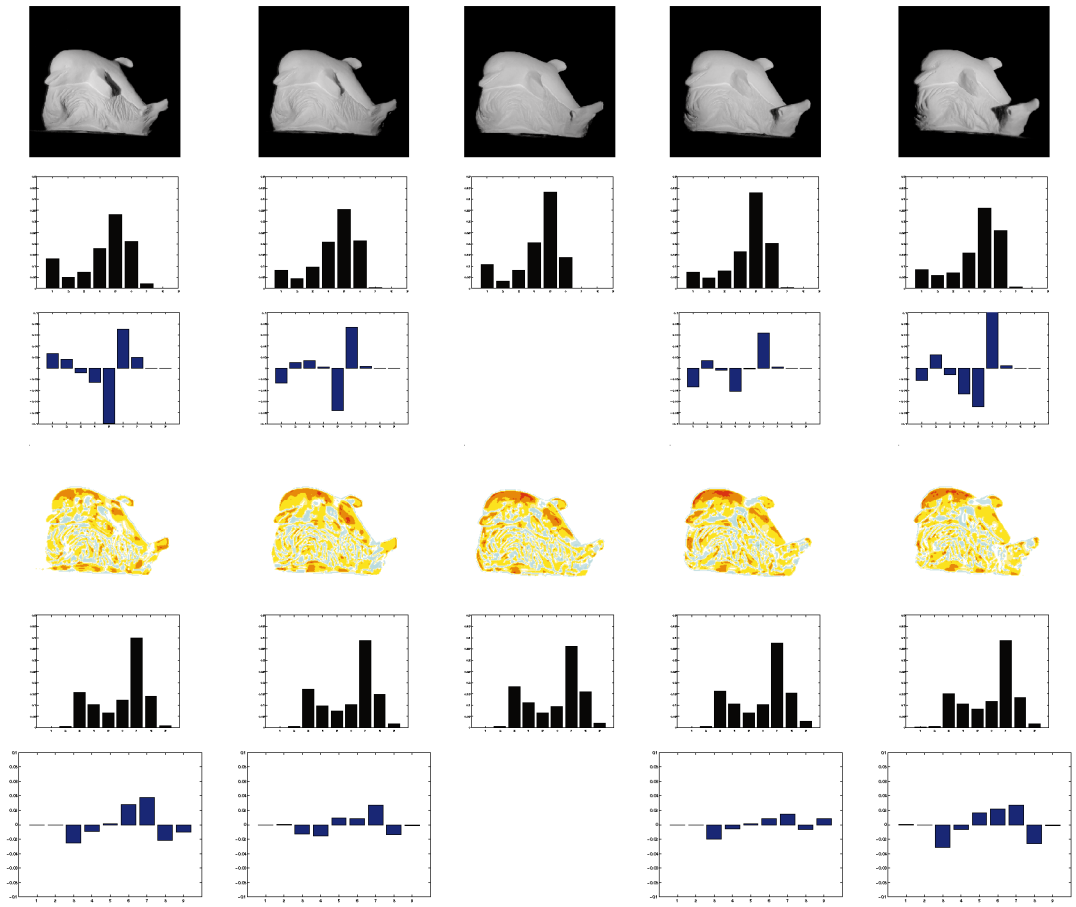


Fig. 3. Labeling an object under variable illumination. Top row, left to right: Images of object illuminated from $-30°$, $-15°$, $0°$, $15°$, and $30°$ to perpendicular in the x-direction. Second row: Corresponding gray-level histograms. Third row: Difference of gray-level histograms from the histogram obtained using a light-source perpendicular to the image plane, i.e., difference from the middle histogram of the second row. Fourth row: Shape-index labelings from SFS. Fifth-row: One-dimensional histograms of shape-index values. Bottom row: Difference of shape-index histograms from the middle histogram of the fifth row, shown to the same scale as the difference plots in the third row.

5.2.2 *Variable-Illumination*

We continue by providing an investigation of the effect of variable illumination on the shape-index information returned by our new SFS scheme.

We aim to show that it is possible to use SFS to recognize objects under variable illumination, without needing to store model views of the object under many different lighting conditions. It should be sufficient to store the shape-index information for a single

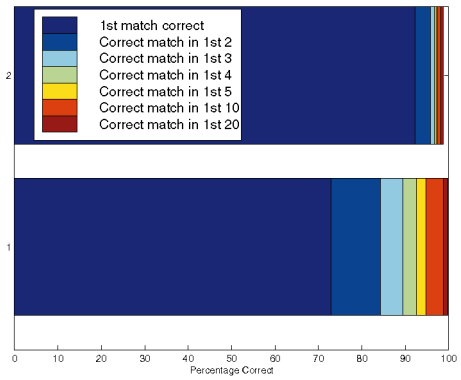


Fig. 4. Recognition results taking the first n closest matches to the query. Bar 1 indicates the results for shape-index histogram matching, using the distribution of local shape index recovered using our curvature consistency SFS scheme, while bar 2 shows recognition using simple gray-level histograms. The shape-index histograms achieve 99.8 percent recognition if the first 20 matches (from 1,160 queries) are considered. This is in spite of the fact that several of the objects in the database are not amenable to analysis by SFS.

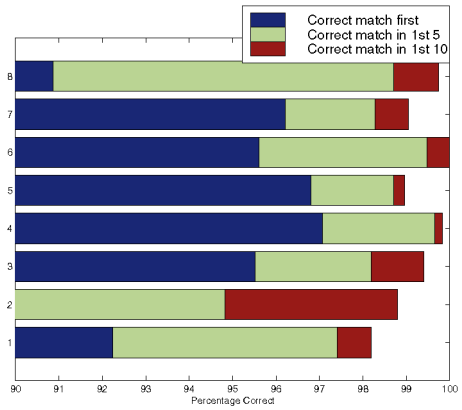


Fig. 5. Recognition results for a variety of histogram representations, taking the first n closest matches to the query. From bottom to top, the bars show the results achievable using: 1) 1D gray level histograms, 2) 1D shape-index histograms, 3) 2D histograms of the azimuthal and inclinational angles of the needle-map, 4) 2D histograms of shape-index and curvedness values, 5) 2D histograms of the principal curvatures (κ_1, κ_2), 6) 2D histograms of Gaussian and Mean curvature, 7) 2D histograms of relative gray-level values, and 8) 2D histograms of relative shape-index values. For the definitions of these histograms, see text. Nine histogram bins are used for the 1D histograms and 9×9 bins for the 2D histograms.

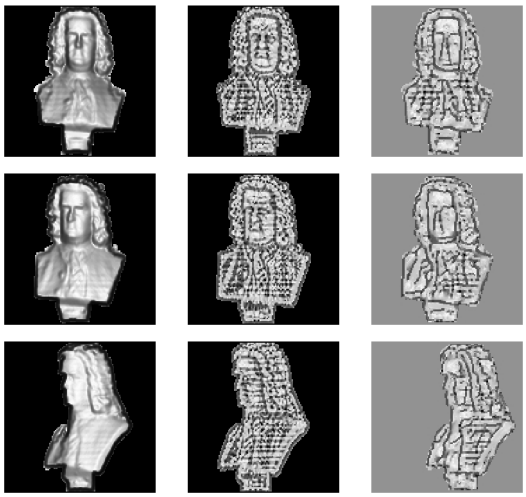


Fig. 6. Examples of the synthetic composer images. The left hand column shows each of the range images illuminated by a light-source coincident with the viewing direction. The middle column shows the shape-index labels extracted directly from the smoothed range images. Finally, the right hand column shows the shape-index labels recovered by applying SFS to the images from the left hand column.

representative view and to use this for recognition under variable illumination and viewpoint.

In practice, the main obstacle to this approach is the problem of shadowing. If the light source is from an oblique angle, large portions of the object may not be illuminated. A further practical obstacle is presented by the limited public availability of databases providing large numbers of views of piecewise-smooth objects under variable illumination conditions. Hence, we resort to using synthetic images generated from range-images. We use seven range images of busts of famous composers obtained from the University of Southern California. Three of the range images are of Bach, two of Brahms, and one each of Beethoven and Chopin. Since there are scanning effects present in the range image data, we apply Gaussian smoothing to the range images prior to illuminating them using the Lambertian model. Fig. 6 illustrates this for the images generated using illumination perpendicular to the image plane (i.e., coincident with the viewing direction). We see that the ground-truth shape-index, generated directly from the smoothed range-image, is extremely noisy. The shape index recovered using SFS draws out salient structures of the images relatively well, but retains significant quantities of scanner-introduced noise which may adversely affect recognition performance by dominating the histograms.

The illumination direction, measured from the perpendicular to the image plane, is taken to have two components, (θ_x, θ_y). These measure the deviation from the perpendicular in the x- and y-axis directions, respectively. To generate the database, θ_x and θ_y are both made to vary over the range $[-26.6°, 26.6°]$. Hence, the most extreme illumination conditions occur when the light source direction is diagonal ($(\theta_x, \theta_y) = \pm 26.6°$) with respect to the image, making an angle of $35.3°$ to the perpendicular.

In spite of the presence of scanner effects, we proceed to attempt object recognition on the basis of SFS-derived representations. The seven images generated by illuminating each range image with a light source in the viewing direction and the corresponding SFS-derived representations are taken as the model images. A further 560 images, generated by illuminating the seven range images from other directions as described above, are taken as queries. Fig. 7a shows the average recognition results achievable using four different histogram types, calculated using all 560 images as queries in turn. In each case, we take the closest three matches. The low recognition rates reflect the difficulty of the task. Specifically, the different objects are very similar, while the effects of the variable illumination upon the images of a given object are relatively large.

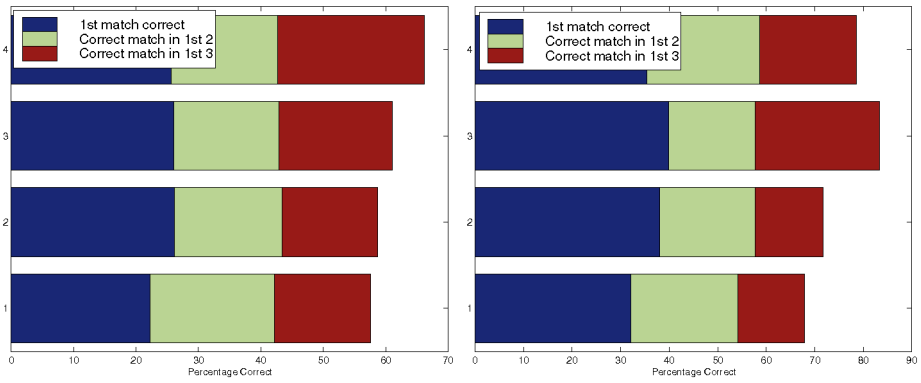
If we consider a restricted range of illumination directions, it is not surprising that the recognition performance improves significantly. Fig. 7b shows the results possible using the restricted lighting ranges $(\theta_x, \theta_y) = [-16.7°, 16.7°]$. Not only does the recognition rate improve for all the schemes, but the advantage of the SFS-derived representations is increased. We take this as evidence that the dominance of shadows noted previously makes a significant contribution to the failures of the SFS-derived representations, although the scanner artifacts present in the data also affects the performance.

5.3 Structural Object Recognition

In this section, we experiment with the recognition strategies described in Section 6. We again use the COIL database from Columbia University in our experiments.

5.3.1 *Individual Recognition Performance*

We commence our study by comparing the performance of each of the different recognition schemes developed in Section 6 in isolation from one another. The results of our experiments are



(a) (b)

Fig. 7. (a) Full set of light source directions. (b) Restricted set of light souce directions. Recognition results obtained on the "Composer" database using 1) 1D gray-level histograms, 2) 1D shape index histograms, 3) 2D histograms of shape-index and curvedness, and 4) 2D histograms of Gaussian and Mean curvature. The first three closest matches to each query are considered.

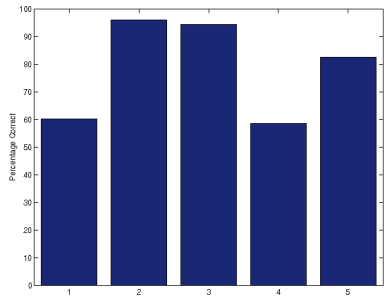


Fig. 8. Comparison of the recognition performance obtained using each component of the structural representation in isolation. The percentage of correct recognitions in the first five matches is shown using 1) the shape-index sequence, 2) curvedness sequence, 3) region size sequence, 4) mean region normals, and 5) the region adjacency graph.

summarized in Fig. 8. The best performance is obtained using the curvedness sequence (98 percent) and the region size sequence (97 percent). The graph-matching method gives a recognition rate of 84 percent. The shape-index sequence (60 percent) and the surface normals (59 percent) give rather disappointing results.

It is interesting to compare the recognition performance of the region-size sequence (97 percent), the region adjacency graph (84 percent), and the shape-index histogram (73 percent). All three methods use information about the relative size of different shape-index regions. The bin-contents of the histogram encode the total region area associated with each shape-index label, but overlook any information concerning relational arrangement. The region-size sequence uses a string ordering of individual region sizes, while the region adjacency graph uses information concerning the relational arrangement of the regions. These results would seem to indicate that, although region-structure is an important source of additional information, the relational arrangement of the regions plays a less important role.

Since the shape-index sequence (60 percent) performs rather poorly, it may be attributed to the fact that it provides little additional information once the regions have been sorted according to size. The reason for this is that the CSMP region segmentation is itself derived from the shape-index. Moreover, the curvedness and the size of the CSMPs are also strongly correlated to one another. Since curvature is inversely proportional to radius, highly curved objects are likely to present a small area. Finally, it is disappointing that the graph-matching method performs less well than the curvedness sequence. However, it is important to note that the sequence is a relational structure. It is a string, where the adjacency relation is the size ordering of CSMPs. The region adjacency graph, on the other hand, uses spatial adjacency as the predicating relation. Hence, the results may be indicative that feature contrast may be more important than spatial organisation. This is certainly in tune with work in the psychology literature including that of Tversky [22].

5.3.2 Combining Evidence

Having considered the individual performance of each of our recognition schemes in turn, we now consider how to combine evidence from different components of the overall representation. We use a simple majority voting procedure. For each recognition scheme in turn, we record the identities of the 10 best matches. These 10 matches each represent a vote that can be cast by a single recognition scheme for the different objects in the database. The different recognition schemes are then polled by tallying their votes for the different objects in the database. The object that receives the greatest number of votes is the winner. The maximum number of votes that a single object can receive is $10e$, where e is the number of recognition algorithms being polled. Clearly, there are many ways of polling committees of experts and this remains an active topic of research. Suffice to say that the aim here is to investigate whether polling can improve the recognition performance significantly.

Fig. 9a shows the recognition results obtained using this simple majority voting approach to combining evidence. We achieve better than 90 percent recognition by considering the first three matches. The results may be slightly improved, as Fig. 9b illustrates, by using only the best two components of the representation as determined from Fig. 8.

6 Conclusions

We have investigated the feasibility of 3D object recognition from 2D images using topographic information derived from SFS. Our contributions are two-fold. First, we have presented an experimental study to investigate the suitability of SFS for simple histogram-based object recognition. The main conclusions are that various curvature-based attributes are sufficiently stable to changes of viewpoint and lighting to be useful for object recognition. Second, we have used CSMPs as a structural abstraction of surface topography. We have also investigated the use of various attributes and relational structures computed from the CSMPs. The most effective of these is a string of curvedness attributes ordered according to CSPM size. However, we have also obtained useful results by matching the region adjacency graph for the CSMPs.

Based on these results, there is clearly a great deal of research that can be undertaken with the aim of improving recognition performance. For example, some form of region compactness or moment-based shape measure or the length of shared perimeter between adjacent regions could also be used for recognition purposes. However, it may prove that the most important contribution of SFS-derived representations has nothing to do with regions. The parabolic lines recovered in the shape-index

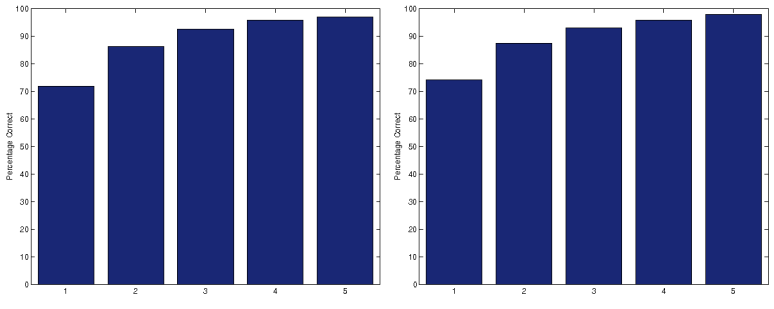


Fig. 9. Overall recognition performance as a function of the number of best matches retained. (a) Individual recognition performance. (b) Combining evidence.

labelings have potential, according to psychophysical observations by Koenderink and van Doorn [17], as a sparse object representation for object recognition.

REFERENCES

[1] P.N. Belhumeur and D.J. Kriegman, "What Is the Set of Images of an Object under All Possible Lighting Conditions?" *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 270-277, 1996.

[2] M. Bichsel and A.P. Pentland, "A Simple Algorithm for Shape from Shading," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 459-465, 1992.

[3] A.M. Bruckstein, "On Shape from Shading," *Computer Vision, Graphics, and Image Processing*, vol. 44, pp. 139-154, 1988.

[4] C. Dorai and A.K. Jain, "COSMOS: A Representation Scheme for 3D Free-Form Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1115-1130, Oct. 1997.

[5] C. Dorai and A.K. Jain, "Shape Spectrum Based View Grouping and Matching of 3D Free-Form Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1139-1146, Oct. 1997.

[6] G.L. Gimelfarb and A.K. Jain, "On Retrieving Textured Images from an Image Database," *Pattern Recognition*, vol. 29, pp. 1461-1483, 1996.

[7] B.K.P. Horn and M.J. Brooks, "The Variational Approach to Shape from Shading," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 2, pp. 174-208, 1986.

[8] B. Huet and E.R. Hancock, "Line Pattern Retrieval Using Relational Histograms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1363-1370, Dec. 1999.

[9] B. Huet and E.R. Hancock, "Fuzzy Relational Distance for Large-scale Object Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 138-143, 1998.

[10] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, Sept. 1993.

[11] A.K. Jain and P.J. Flynn, *Three Dimensional Object Recognition Systems*. Amsterdam: Elsevier, 1993.

[12] R. Kimmel and A.M. Bruckstein, "Tracking Level-Sets by Level-Sets: A Method for Solving the Shape from Shading Problem," *Computer Vision and Image Understanding*, vol. 62, no. 1, pp. 47-58, 1995.

[13] R. Kimmel, K. Siddiqi, B.B. Kimia, and A.M. Bruckstein, "Shape from Shading—Level Set Propagation and Viscosity Solutions," *Int'l J. Computer Vision*, vol. 16, no, 2, pp. 107-133 1995.

[14] J.J. Koenderink, *Solid Shape*. Cambridge Mass.: MIT Press, 1990.

[15] J.J. Koenderink and A.J. van Doorn, "Surface Shape and Curvature Scales," *Image and Vision Computing*, vol. 10, pp. 557-565, 1992.

[16] J.J. Koenderink, A.J. van Doorn, and A.M.L. Kappers, "Surface Perception in Pictures," *Perception and Psychophysics*, vol. 52, no. 5, pp. 487-496, 1992.

[17] J.J. Koenderink and A.J. van Doorn, "The Internal Representation of Solid Shape with Respect to Vision," *Biological Cybernetics*, vol. 32, pp. 211-216, 1979.

[18] E. Mingolla and J.T. Todd, "Perception of Solid Shape from Shading," *Biological Cybernetics*, vol. 53, pp. 137-151, 1986.

[19] J. Oliensis and P. Dupuis, "An Optimal Control Formulation and Related Numerical Methods for a Problem in Shape Reconstruction," *Ann. Application Probability*, vol. 4, no. 2, pp. 287-346, 1994.

[20] M.J. Swain and D.H. Ballard, "Color Indexing," *Int'l J. Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

[21] E. Rouy and A. Tourin, "A Viscosity Solutions Approach to Shape-from-Shading," *SIAM J. Numerical Analysis*, vol. 29, no. 3, pp. 867-884, 1992.

[22] A. Tversky, "Features of Similarity" *Psychological Rev.*, vol. 84, pp. 327-352, 1977.

[23] P.L. Worthington and E.R. Hancock, "Histogram-Based Object Recognition Using Shape-from-Shading," *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, pp. 643-648, 2000.

[24] P.L. Worthington and E.R. Hancock, "Region-Based Object Recognition Using Shape-from-Shading," *Proc. European Conf. Computer Vision*, pp. 455-471, 2000.

[25] P.L. Worthington and E.R. Hancock, "Needle Map Recovery Using Robust Regularizers," *Image and Vision Computing*, vol. 17, no. 8, pp. 545-559, 1998.

[26] P.L. Worthington and E.R. Hancock, "3D Surface Topography from Intensity Images," *Proc. IEEE Int'l Conf. Computer Vision*, vol. II, pp. 911-917, 1999.

[27] P.L. Worthington and E.R. Hancock, "New Constraints on Data-Closeness and Needle-Map Consistency for SFS," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1250-1267, Dec. 1999.

[28] P.L. Worthington, "Enhanced Shape-from-Shading for Object Recognition," PhD thesis, Univ. of York, 1999.