Increased Extent of Characteristic Views using Shape-from-Shading for Object Recognition

Philip L. Worthington Benoit Huet Edwin R. Hancock
Department of Computer Science, University of York, UK
[plw|huetb|erh]@minster.cs.york.ac.uk

Abstract

This paper investigates the use of shape-from-shading for object recognition. The local surface orientation information recovered using shape-from-shading is shown to provide useful input to an appearance-based object recognition scheme. We consider two representations which may be recovered from shading information - the needle-map, and the local curvature shape-index - and examine their relative performance for object recognition. Specifically, we use a histogram-comparison technique, and focus upon the relative stability of the representations to small changes of viewpoint. We demonstrate that the needle-map representation allows the view-sphere to be spanned using a significantly smaller number of characteristic views than using either the raw images or the shape index.

1 Introduction

Despite long-term interest in shape-from-hading (SFS), and psychophysical evidence that it is a key process in 3D surface perception [15], there are few reports of its use in practical object-recognition systems [27]. One of the principal reasons for this is the lack of robust algorithms capable of recovering fine surface detail. Instead, much of the effort in the literature has focused on appearance-based object recognition using either iconic [18] or grey-scale manifolds [16]. This is a disappointing omission, since SFS can provide direct information concerning surface topography, for example characteristic, or typical, views [22, 19] and aspect graphs [10, 21].

View-based representations have recently been demonstrated to provide a powerful means of recognising 3D objects [20, 4, 12, 17, 24]. In essence the technique relies on constructing a distributed 3D representation which consists of a series of characteristic or typical 2D views. For instance, Seibert and Waxman [20] have a Hough-like method in which different views form distinct clusters in accumulator space. Gigus and Malik [4] present a method for computing the aspect graphs of polyhedra in line-drawings using visual events for faces, edges and vertices. Kriegman [12] uses the algebraic structure of occluding contours, whilst Petitjean [17] has developed these ideas to extract visual event surfaces for piecewise smooth objects. Several authors have considered the statistical distribution of characteristic views. For instance Malik and Whangbo [14] have shown that it is inappropriate to distribute the nodes of the aspect graph uniformly across the view-sphere. In a similar vein, Weinshall and Werman have characterised both the likelihood and stability of different characteristic views [24]. These ideas have been applied to

BMVC 1998 doi:10.5244/C.12.71

the recognition of objects from large model-bases [23]. Meanwhile, Dorai and Jain have recently shown how histograms of surface curvature attributes can be used to recognise different views of curved objects in range images [3].

In practice, view-based object recognition is most easily realized if the different views are organised using either a geometric or relational structure. An example of the former is the view-sphere, while the latter is typified by the aspect graph. Although offering a convenient view-based object representations, both the view-sphere and the aspect graph have proved to be notoriously difficult to elicit from real-world imagery.

Our aim here is to consider how SFS can be used to generate a view-based representation of object appearance, and how this can in turn be used for 3-D object recognition using 2-D views. The starting point for our study is a recent series of papers [26, 25] in which we have reported an improved shape-from-shading algorithm using robust-regularizers. The main advantage of this method is to limit the over-smoothing of fine curvature detail. The main contribution is to investigate whether needle-maps can be used for 3D object recognition. We develop two alternative, histogram-based recognition strategies, the first using the surface normals directly, and the second based upon the shape index of Koenderink and van Doorn [11].

The recognition strategies are evaluated on the Columbia University data-base of 20 arbitrarily-selected, real-world objects. Here we show that both representations provide useful recognition performance. However, the surface-normal histogram is found to be more effective than the shape-index histogram. A sensitivity study reveals that the method offers significant discrimination to the differential topology of object appearance on the view sphere. In other words, our needle-maps provide a viable computational basis for automatically extracting characteristic views from 2D images of 3D objects.

2 Shape from Shading

Shape-from-shading (SFS) has been an active subject of research for over two decades, and may be regarded as one of the classical problems of computer vision. In recent research we have developed a SFS technique based upon the variational approach of Horn and Brooks [1, 7, 8]. Our scheme addresses one of the main problems with the Horn and Brooks technique - its tendency to over-smooth the recovered needle-map, leading to a loss of detail in regions where the surface orientation varies rapidly. Several other solutions have been proposed to this (e.g. [6]), but our research has shown that the apparatus of robust statistics may be applied to the problem with encouraging results [26, 25].

In brief, we wish to solve the normalized image irradiance equation

$$E(x, y) = R(p, q) \tag{1}$$

where $E(x, y)$ is the image of the object, and $R(p, q)$ is the reflectance of a surface patch oriented such that its normal has direction $\mathbf{n} = (-p, -q, 1)^T$. The quantities p and q are the components of the surface gradient in the x and y direction respectively, i.e. $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$.

If the surface is assumed to have Lambertian reflectance properties, the brightness of a patch will simply be proportional to the angle between the surface normal and the light source direction, $\mathbf{s}$. The image irradiance equation then becomes $E(x, y) = \mathbf{n} \cdot \mathbf{s}$. Unfortunately, this is under-constrained for the recovery of p and q over most of an

object's surface. Hence, we must introduce an additional constraint on the smoothness of the recovered needle-map. This is encoded by constructing an energy functional of the form

$$I = \int\int \Big(E(x,y) - \mathbf{n}\cdot\mathbf{s}\Big)^2 + \lambda\left(\rho_\sigma\left(\left\|\frac{\partial\mathbf{n}}{\partial x}\right\|\right) + \rho_\sigma\left(\left\|\frac{\partial\mathbf{n}}{\partial y}\right\|\right)\right)dxdy \qquad (2)$$

where ρ_σ may be any regularization function, and λ is a Lagrange multiplier. The first term of this functional encodes the image irradiance equation. The second term uses the derivatives of the recovered normals to penalize sharp changes of orientation according to the function ρ_σ.

Applying the calculus of variations and discretizing the resulting Euler equation, we develop the following generalized update equation for iteratively estimating the surface normals

$$\begin{aligned}
\mathbf{n}_{i,j}^{(k+1)} &= \Big(E - \mathbf{n}_{i,j}^{(k)}\cdot\mathbf{s}\Big)\mathbf{s} \\
&+ \frac{\lambda}{2}\left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial x}\right\|^{-1}\left[\frac{\partial}{\partial x}\left(\rho'_\sigma\left(\left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial x}\right\|\right)\right) + \rho'_\sigma\left(\left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial x}\right\|\right)\times\right. \\
&\left.\left(\mathbf{n}_{i+1,j}^{(k)} + \mathbf{n}_{i-1,j}^{(k)} - \left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial x}\right\|^{-2}\left(\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial x}\cdot\frac{\partial^2\mathbf{n}_{i,j}^{(k)}}{\partial x^2}\right)\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial x}\right)\right] \\
&+ \frac{\lambda}{2}\left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial y}\right\|^{-1}\left[\frac{\partial}{\partial y}\left(\rho'_\sigma\left(\left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial y}\right\|\right)\right) + \rho'_\sigma\left(\left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial y}\right\|\right)\times\right. \\
&\left.\left(\mathbf{n}_{i,j+1}^{(k)} + \mathbf{n}_{i,j-1}^{(k)} - \left\|\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial y}\right\|^{-2}\left(\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial y}\cdot\frac{\partial^2\mathbf{n}_{i,j}^{(k)}}{\partial y^2}\right)\frac{\partial\mathbf{n}_{i,j}^{(k)}}{\partial y}\right)\right]
\end{aligned}$$

In the quadratic case where $\rho_\sigma(\eta) = \eta^2$, this becomes the update equation used by Horn and Brooks [1]. However, any other function may be used as the regularization term, and we have investigated several robust measures, including the classical Tukey [5] and Huber [9], and the Adaptive Prior Potential Functions of Li [13]. We also introduced [25] a continuous version of the piecewise Huber robust estimator, described by

$$\rho_\sigma(\eta) = \frac{\sigma}{\pi}\log\cosh\left(\frac{\pi\eta}{\sigma}\right) \qquad (3)$$

and found that this yielded the best results by offering a compromise between over-smoothing and noise rejection/numerical stability.

3 Characteristic Views

The concept of a characteristic view (CV) is useful in appearance-based object recognition [22]. It stems from the desire to obtain a *representative and adequate grouping* of views, such that a given level of recognition accuracy may be achieved using the minimum number of stored views [3]. Clearly, this has important implications for the storage space needed to represent each object, and the number of matches which must be performed at

run-time for the purpose of recognition. View grouping has been addressed using CVs and aspect graphs (AG). An aspect graph [10] enumerates all possible appearances of an object, and the change in appearance at the boundary between different aspects is called a visual event.

However, aspect graphs grow to unwieldy sizes for complex, non-polyhedral objects, since all visual events are considered sufficiently important to define a new boundary between aspects[17]. It is difficult to define a single face when an object is composed of piecewise curved surfaces[12]. Even slight changes in viewpoint may result in more of the curved surface(s) either coming into, or disappearing from, the view. Thus, either the size of the aspect graph must be controlled using appropriate heuristics [23], or a less rigid approach considered. We choose to adopt the latter course, and treat the concept of a characteristic view in a more psychophysical manner, as a natural groupings of views.

A possible method of identifying natural CVs, in this sense, is to use clustering to identify natural view groupings [20]. From a human perspective, all views of an object which form a CV should "look" more similar to each other than to any view from a different CV. If all the views within a CV are similar, then only one such view (or an average view) need be stored and matched for recognition. It follows that the larger, on average, each CV is, the fewer model views need be stored in order to span the view-sphere, and the more efficient both the learning and recognition of objects will become.

The representation used for the model views has great influence upon the average extent of the CVs. A representation which is relatively stable over a range of viewpoints will result in larger CVs, on average, than one which changes greatly for small shifts in viewpoint. However, this local invariance must not be at the expense of loss of detail, since this will impair the ability to discriminate between objects.

4 Using SFS for Object Recognition

There are three obvious ways to utilize the orientation information encapsulated by the needle-map. Most of the literature focuses exclusively upon the first of these; the integration of local orientation information to recover an approximation to the object surface [6]. In the context of object recognition, this is most useful for model-based recognition. In practice, however, the accurate and reliable recovery of surfaces through SFS has proved extremely difficult. The second approach is to use the needle-map directly. In other words, instead of storing 2-D model views, we store $2\frac{1}{2}$D models and match on orientation information. A third approach is to calculate a physically meaningful local surface description. An obvious example is local surface curvature.

4.1 Direct Use of Needle-Map

The needle-map is a valid representation for object recognition. In terms of dimensionality of the matching representation, it may be viewed as midway between model (3-D) and appearance-based (2-D) recognition. However, since a series of model needle-maps are needed for each object, it remains essentially an appearance-based technique. If we deal with unit normals, two values are sufficient to describe the direction of each normal, since the third component may be determined from the other two. Thus, matching can be performed using 2-D vectors.

4.2 The Shape Index

The differential structure of a surface is captured by the local Hessian matrix, which may be approximated in terms of surface normals by

$$\mathcal{H} = \begin{pmatrix} -\left(\frac{\partial \mathbf{n}}{\partial x}\right)_x & -\left(\frac{\partial \mathbf{n}}{\partial x}\right)_y \\ -\left(\frac{\partial \mathbf{n}}{\partial y}\right)_x & -\left(\frac{\partial \mathbf{n}}{\partial y}\right)_y \end{pmatrix} \tag{4}$$

where $(\cdots)_x$ and $(\cdots)_y$ denote the x and y components of the parenthesized vector respectively.

The principal curvatures of the surface are the eigenvalues of the Hessian matrix, found by solving $|\mathcal{H} - \kappa\mathbf{I}| = 0$ for κ, where $\mathbf{I}$ is the identity matrix. Koenderink and van Doorn[11] developed a single-value, angular measure to describe local surface topology in terms of the principal curvatures. This *shape index* is defined as

$$s = \frac{2}{\pi} \arctan \frac{\kappa_2 + \kappa_1}{\kappa_2 - \kappa_1} \qquad \kappa_1 \geq \kappa_2 \tag{5}$$

and may be expressed in terms of surface normals thus

$$s = \frac{2}{\pi} \arctan \frac{\left(\frac{\partial \mathbf{n}}{\partial x}\right)_x + \left(\frac{\partial \mathbf{n}}{\partial y}\right)_y}{\sqrt{\left(\left(\frac{\partial \mathbf{n}}{\partial x}\right)_x - \left(\frac{\partial \mathbf{n}}{\partial y}\right)_y\right)^2 + 4\left(\frac{\partial \mathbf{n}}{\partial x}\right)_y \left(\frac{\partial \mathbf{n}}{\partial y}\right)_x}} \tag{6}$$

Figure 1 shows the range of shape index values, the type of curvature which they represent, and the grey-levels used to display different shape-index values. Dark regions correspond to concavities, such as ruts, troughs and spherical caps, whilst light regions indicate caps, domes and ridges.

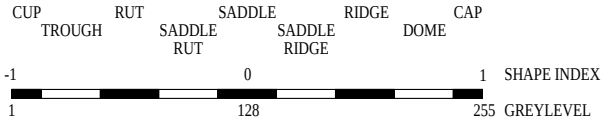

Figure 1: The shape index scale ranges from -1 to 1 as shown. The shape index values are encoded as a continuous range of grey-level values between 1 and 255, with grey-level 0 being reserved for background and flat regions (for which the shape index is undefined).

5 Experiments

To compare the different representations, we use a standard histogram recognition scheme [2]. Although this does not take into account the spatial arrangement of an image, it is useful in identifying CVs of objects, since it gives a good indication of the stability of a representation to small changes of viewpoint. The behaviour of the different measures under the histogram recognition procedure enables qualitative assessment of the representations in terms of average CV extent.

We measure the proximity between two images using the Bhattacharyya distance

$$B(P_Q, P_M) = -\ln \sum_{i=1}^{n} \sqrt{P_Q(i) \times P_M(i)}$$

where P_Q is the query histogram and P_M one of the model histograms.

Figure 2 illustrates the results of our experiments for 4 of the 20 images in the test set. This image set is the Columbia Image Object Library, consisting of 20 arbitrary objects. There are 72 views of each object, illuminated by a light source coincident with the camera. The images are taken at 5° intervals along a great circle of the object's view-sphere. Only around 9% of the view-sphere is spanned by these 72 images, underlining the need for view grouping if appearance-based object recognition is not to require unfeasibly large numbers of models.

The first row of Figure 2 shows the first image from each of the 72 view sequences for 4 objects in the dataase. The second row shows the needle-maps recovered by the SFS technique described in Section 2, whilst the third row displays the shape index classes derived from the needle-map. The grey-levels correspond to the scale in Figure 1.

Rows 4-6 of Figure 2 show the histograms for each of the object representations in turn. In each case, the leftmost bin corresponds to background pixels and is excluded from the calculation of Bhattacharyya distance between the histograms.

Row 4 shows the grey-level histograms for the raw images, and Row 5 the 2-D histograms of the needle-maps. Clearly, there is a great deal of variability in the structure of these 2-D histograms.

The shape-index histograms of Row 6 are all broadly similar. Each is bi-modal, with the two modes corresponding approximately to ruts and ridges/domes.

Figure 3 shows histogram ranking results for each of the representations. These are average plots taken over all 72 images representing a given object. In each case, one of the 72 images is chosen as the query image, and all 1440 images in the database ranked according to their distance from this query. Clearly, the query image itself has zero self-distance and hence is ranked 0. Views of the same object from similar viewpoints, i.e. those with small angular deviations in any direction on the viewsphere, should come next in the ranking, and so on. Each image in the set representing a given object is taken as the query in turn, and an average ranking found for all images at a given angular distance either side of the query. This is repeated for each of the object representations.

To establish CVs, we require a representation which provides a good ranking ability over as wide a range of angular distance as possible. The surface normal representation clearly meets this requirement in each of the cases shown. Specifically, it provides a better ranking ability over a wider range of angular distances than the raw images. The shape-index also does relatively well for the first two objects, but is unstable to even small changes in viewing angle for the second pair of objects. The latter images contain significant surface markings, resulting in rapid changes of albedo. These break the fundamental Lambertian assumptions underlying our SFS technique, leading to poor needle-map recovery in these regions. The recovery errors are subsequently compounded in the calculation of the shape-index.

Figure 4 shows the averaged ranking results, over the full $\pm 180^\circ$ range of angular distances. Here we display the result of taking each of the 1440 images as the query image in turn and averaging the rankings of all images of the same object as the query. The results are plotted as a function of the angular distance from the query. We use only

one bin size for each representation. The shape-index does poorly in comparison to the raw intensity images. However, there is a clear advantage in using the needle-map as the average ranking remains much lower over a wider range of angular distances from the query image.

6 Conclusions and Outlook

We have demonstrated that the needle-map is a useful representation for object recognition, proving more stable to small changes of viewpoint than raw intensity images. This implies a significant saving in the number of model views which must be stored and matched for each object.

We have also investigated the use of the shape index, a measure designed to capture variations of surface curvature. Dorai and Jain[3] have recently reported excellent results using this physically-motivated measure with range images, once again enabling significant grouping into CVs of an object to occur. However, in conjunction with SFS, the shape index performs significantly worse than using the needle-map directly.

There is extensive scope for further work, not least because the results presented here are derived using an extremely simple recognition technique. A more rigorous analysis is needed of how many CVs need be stored to achieve the same recognition accuracy using the needle-map and the raw image representations.

References

[1] M.J. Brooks and B.K.P. Horn. Shape and source from shading. *IJCAI*, pages 932–936, 1986.

[2] P.A. Devijver and J. Kittler. *Pattern Recognition-A Statistical Approach*. Prentice-Hall, 1982.

[3] C. Dorai and A.K. Jain. Shape spectrum based view grouping and matching of 3d free-form objects. *IEEE PAMI*, 19(10):1139–1146, 1997.

[4] Z. Gigus and J. Malik. Computing the aspect graph for line drawings of polyhedral objects. *IEEE PAMI*, 12(2):113–122, 1990.

[5] D.C. Hoaglin, F. Mosteller, and J.W. Tukey. *Understanding robust and exploratory data analysis*. Wiley, New York, 1983.

[6] B.K.P. Horn. Height and gradient from shading. *IJCV*, 5(1):37–75, 1990.

[7] B.K.P. Horn and M.J. Brooks. The variational approach to shape from shading. *CVGIP*, 33(2):174–208, 1986.

[8] B.K.P. Horn and M.J.(eds) Brooks. *Shape from Shading*. MIT Press, Cambridge, MA, 1989.

[9] P. Huber. *Robust Statistics*. Wiley, Chichester, 1981.

[10] J.J. Koenderink and A.J. van Doorn. The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32:211–216, 1979.

[11] J.J. Koenderink and A.J. van Doorn. Surface shape and curvature scales. *IVC*, 10:557–565, 1992.

[12] D.J. Kriegman. Computing stable poses of piecewise smooth objects. *Computer Vision, Graphics and Image Processing*, 55(2):109–118, 1992.

[13] S.Z. Li. Discontinuous mrf prior and robust statistics: a comparative study. *IVC*, 13(3):227–233, 1995.

[14] R. Malik and T. Whangbo. Angle densities and recognition of 3d objects. *IEEE PAMI*, 19(1):52–57, 1997.

[15] D.C. Marr. *Vision*. Freeman, San Francisco, 1982.

[16] S.K. Nayar, H. Murase, and S.A. Nene. Parametric appearance representation. *in Early Visual Learning, Oxford University Press*, 1996.

[17] S Petitjean. The enumerative geometry of projective algebraic-surfaces and the complexity of aspect graphs. *IJCV*, 19(3):261–287, 1996.

[18] R.P.N. Rao and D.H. Ballard. An active vision architecture based on iconic representations. *AI*, 78:461–505, 1995.

[19] J. Rieger. The geometry of view space of opaque objects bounded by smooth surfaces. *AI*, 44:1–40, 1990.

[20] M. Seibert and A.M. Waxman. Adaptive 3-d object recognition from multiple views. *IEEE PAMI*, 14(2):107–124, 1992.

[21] J.H. Stewman and K.W. Bowyer. Aspect graphs for convex planar-face objects. *Proc. IEEE Workshop on Computer Vision*, pages 123–130, 1987.

[22] R. Wang and H. Freeman. Object recognition based on characteristic view classes. *Proc. ICPR*, I:8–12, 1990.

[23] D. Weinshall and M. Werman. Disambiguation techniques for recognition in large databases and for under-constrained reconstruction. *Proc. IEEE Symposium on Computer Vision*, pages 425–430, 1995.

[24] D. Weinshall and M. Werman. On view likelihood and stability. *IEEE PAMI*, 19(2):97–108, 1997.

[25] P.L. Worthington and E.R. Hancock. Needle map recovery using robust regularizers. *Proc. British Machine Vision Conference*, I:31–40, 1997.

[26] P.L. Worthington and E.R. Hancock. Shape-from-shading using robust statistics. *Proc. IEEE Int. Conf. on Digital Signal Processing*, 1997.

[27] A.L. Yuille, M. Ferraro, and T. Zhang. Surface shape from warping. *Proc. CVPR*, pages 846–851, 1997.

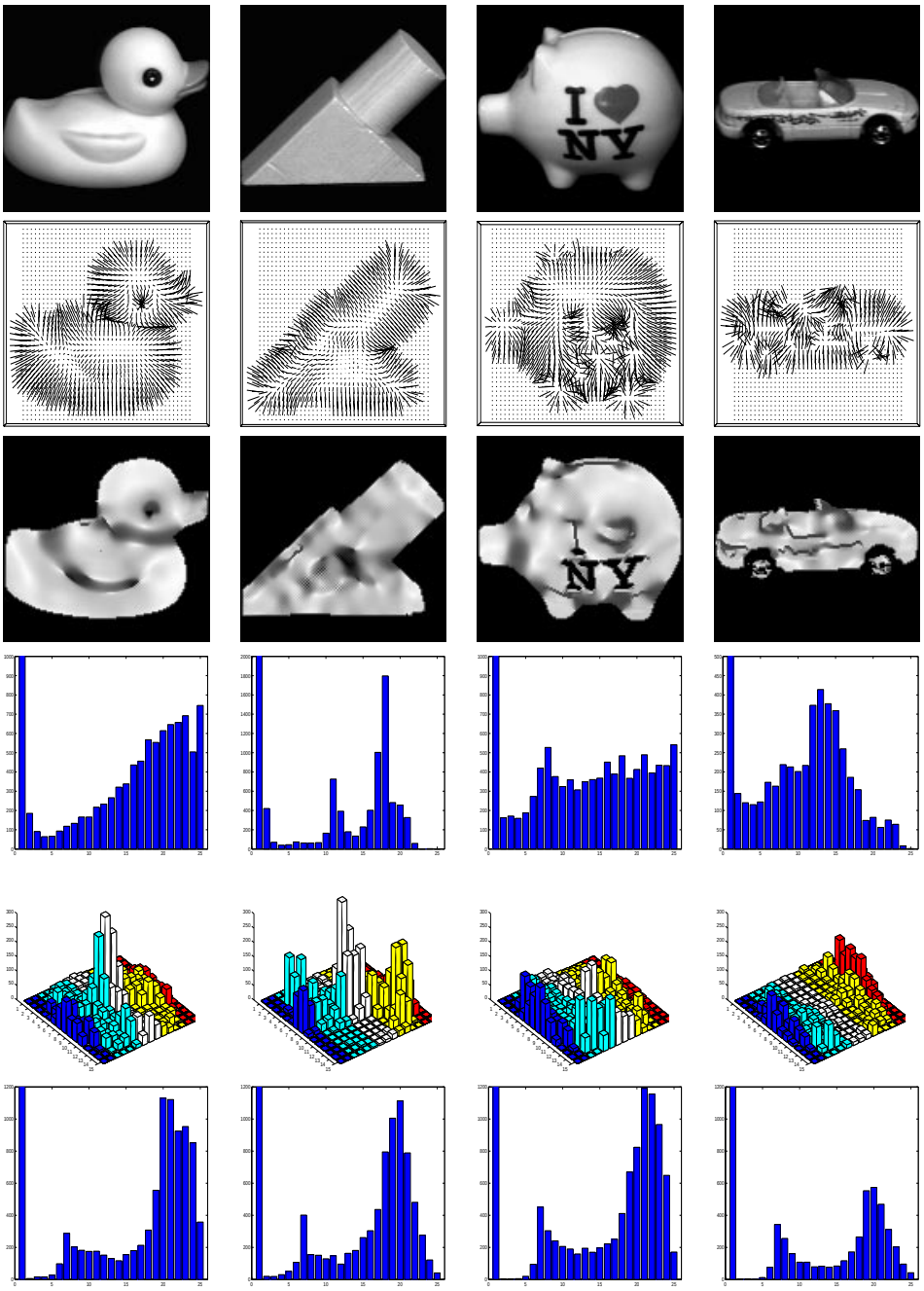

Figure 2: Top row: Raw Images. Row 2: Recovered Needle-maps. Row 3: Shape Index representation. Row 4: 25 bin grey-level frequency histograms. Row 5: 15x15 bin 2-D histograms of normal direction frequency. Row 6: 25 bin shape index frequency histograms.

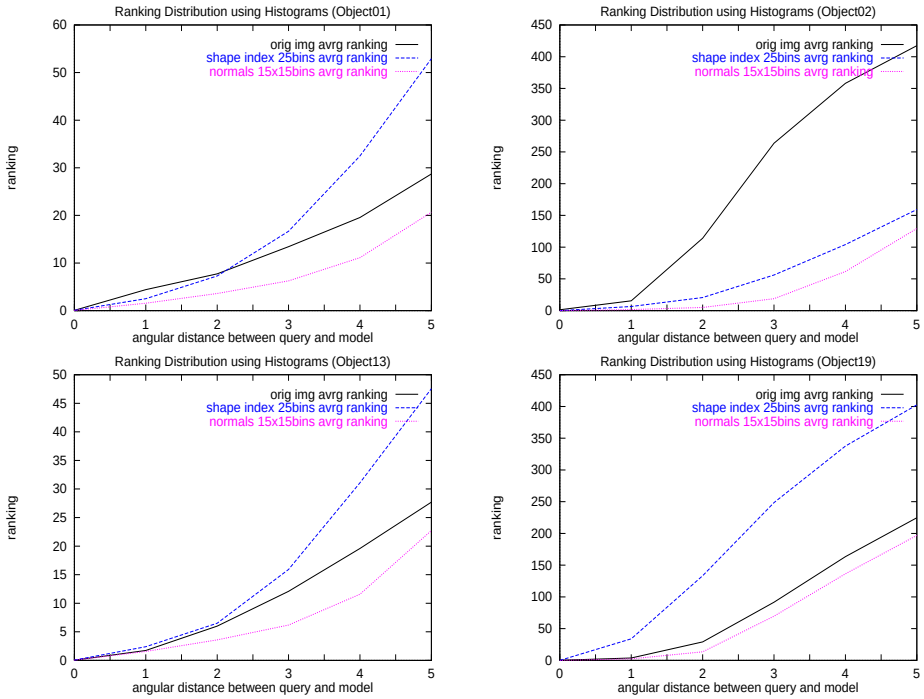

Figure 3: Plots of average ranking vs distance from query over all images of a given object. Each one of the 72 images of the object is taken as the query image in turn, and all 1440 images in the database ranked according to their distance from this query. The average ranking found for all images at a given angular distance either side of the query. An angular distance of 1 represents the average of the images at $\pm 5^\circ$ from the query.

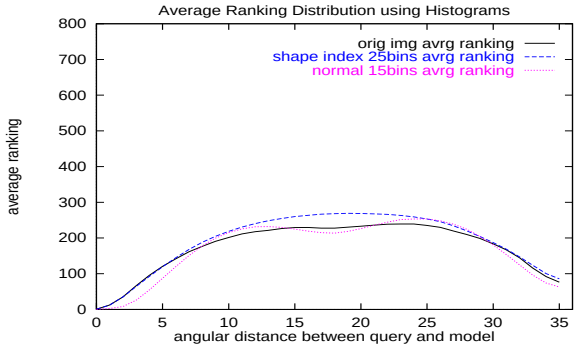

Figure 4: Plots of average ranking vs distance from query over all images in the database. Each of the 1440 images is taken as the query image in turn. The dip around angular distance 18 ($\pm 90^\circ$), and the larger dip towards angular distance 35 ($\pm 180^\circ$), are attributable to a number of the objects possessing approximate rotational symmetry of order 2 and 4 respectively.