Face Recognition with Region Division and Spin Images

Yang Li, William A.P. Smith, and Edwin R. Hancock

Department of Computer Science, University of York, York, YO10 5DD, UK

Abstract. This paper explores how spin images can be constructed using shape-from-shading information and used for the purposes of face recognition. We commence by extracting needle-maps from gray-scale images of faces, using a mean needle-map to enforce the correct pattern of facial convexity and concavity. Spin images [4] are estimated from the needle maps using local spherical geometry to approximate the facial surface. Our representation is based on the spin image histograms for an arrangement of image patches. We demonstrate how this representation can be used to perform face recognition across different subjects and illumination conditions. Experiments show the method to be reliable and accurate, and the recognition precision reaches 98% on CMU PIE sub- database.

1 Introduction

Face recognition is an active research area that has been approached in many ways. Roughly speaking the alternative methods can be divided in two categories. The first of these is the feature-based method, while the second is the model-based method. Recently, it is the model-based method that has attracted the greatest attention [2]. Here one on the most important recent developments is the work of Blanz and Vetter [3]. In this work a 3-D morphable model matched to face data using correspondences delivered by optic flow information. The method gives recognition rates of about 80% when profiles are used to recognise frontal poses. However, the construction of the model requires manually marking feature points, which is labour intensive. Hence, the automatic construction of models remains an imperative in face recognition. There are related feature-based approaches which are based on the assumption that face images are the result of Lambertian reflectance. Under this assumption 3D linear subspaces can be constructed that account for facial appearance under fixed viewpoint but under different illumination [9,1,7].

In this paper we aim to develop a feature based method for face recognition that can be used to recognise faces using surface shape information inferred from image brightness using a Lambertian shape-from-shading scheme. Shape-from-shading is not widely accepted as a technique for face recognition. The reason is that the surface normal is commonly believed to be noisy and is unstable under changes in illumination direction or change of pose. However, recently it has been shown that shape-from-shading can be used to extract useful features from real world face images [8].

One of the problems that hinders the extraction of reliable facial topography using shape-from-shading is the concave/convex inversions that arise due to the bas-relief ambiguity. A recent paper [8] have shown how this problem can be overcome using a statistical model for admissible surface normal variations trained on range data. Here

we use a simplified version of this algorithm. The surface normals are constrained to fall on the Lambertian reflectance with axis in the light source direction and apex angle given by the inverse cosine of the normalised image brightness. The position of the surface normal on the cone is such that it minimizes the distance to the corresponding mean surface normal direction.

To construct a surface representation from the surface normals, we turn to the spin image first developed by Johnson and Hebert [4]. A spin image is a series of histograms constructed from the polar coordinates of arbitrary reference points on a surface. The representation can capture fine topographic surface detail. Unfortunately, the computational overheads associated with the method are high, since a histogram needs to be generated for each surface location. Moreover, the original spin image representation was developed for range images and hence relies on surface height rather than surface normal information. We demonstrate how these two problems can be overcome by computing local spin images on image patches using surface normal information.

2   Mean Needle Map Alignment

The shape-from-shading algorithm used to extract needle-maps from brightness images is as follows. We follow the work in [10] and place the surface normal on a cone whose axis is the light source direction and whose opening angle is the inverse cosine of the normalised image brightness.

This initial field of surface normals typically contains errors, and in particular locations where the pattern of convexity or concavity is reversed. To overcome this problem we draw on a model that accounts for the distribution of surface normals across ground-truth facial surfaces. To construct this model we use a sample of range images of human faces. From the gradients of the surface height data, we make estimates of surface normal direction. The resulting fields of surface normals are adjusted so that that the faces have the same overall centering, scale and orientation. At each location we compute the mean surface normal direction over the set of training images. Here we use the Max-Planck data-base which has 200 sample images of male and female subjects.

We use the mean facial needle-map to adjust the positions of the surface normals on the reflectance cones. Each initial surface normal is rotated on its cone so that it minimises the angle subtended with the mean surface normal at the corresponding image location.

$$f(x, y) = argmin(tan(\theta_r(x, y) - \theta_{mean}(x, y))) \quad (1)$$

where θ_r and θ_{mean} are the azimuth angles of the aligned surface normal n_r and the mean surface normal n_{mean} on the surface point (x, y).

The simplest way to satisfy 1 is to adjust the azimuth angle of the aligned surface normal n_r so that n_r becomes parallel to n_{mean}.

In Fig. 3 we illustrate the improvements gained using this simple shape-from-shading procedure. In the top row of the figure we show the input images of a single subject with the light source in different directions. In the second row we show the initial estimates of the surface normal directions. Here we have visualised the needle-maps by taking the inner product of the surface normal with the light-source vector perpendicular to

the image plane. This is equivalent to re-illuminating the field of surface normals with frontal Lambertian reflectance. From the images in the second row it is clear that there are significant concave/convex inversions in the proximity of the nose and lips when the face is illuminated obliquely. In the third row of the figure we show the field of surface normals that result from the adjustment procedure described above. The re-illuminations reveal that the inversions are removed and the quality of the recovered facial topography is improved (Fig. 4 illustrates the solution of this inversion problem).

3   The Spin Image Approach

The spin image of Johnson and Hebert [4] aims to construct an object-centered representation. The representation consists of a series of 2D histograms and is constructed in the following manner: Commence by selecting an arbitrary point on the surface as the reference point O, and $\vec{n_o}$ is the surface normal at the point O. Then select a second arbitrary point P on the surface, and $\vec{n_p}$ is the surface normal at the point P. Assume the object resides in a 3D coordinate system with the surface normal $\vec{n_o}$ as z axis and the xy plane perpendicular to $\vec{n_o}$. The Euclidean distance $\gamma = |\vec{OP}|$ can be projected onto the xy plane as α and the z axis as β respectively. After the distances α and β of all the surface points are calculated, we can use them to construct a 2D histogram. The above procedure is performed after each point on the surface has been taken as the point P so that a single 2D histogram is constructed, and then a series of 2D histograms are constructed using the above steps and taking each point on the surface as the reference point O. Figure 1 illustrates the spin image construction. Equation 2 shows the structure of the spin image, in which N is the number of image surface points and n is the number of histogram bins.

$$
\begin{matrix}
O_1 & O_2 \\
\begin{bmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
a_{21} & a_{22} & \dots & a_{2n} \\
\vdots & & \vdots & \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{bmatrix}
&
\begin{bmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
a_{21} & a_{22} & \dots & a_{2n} \\
\vdots & & \vdots & \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{bmatrix} \\
\dots & O_N \\
& \begin{bmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
a_{21} & a_{22} & \dots & a_{2n} \\
\vdots & & \vdots & \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{bmatrix}
\end{matrix}
\tag{2}
$$

This object-centered representation is invariant to translation and rotation since the spin image is calculated using only relative distances between object surface points.

The spin image representation is based on the availability of surface height data and can not be applied directly to fields of surface normals or needle maps. Moreover, the spin image histograms need to be constructed at each image location, and this is computationally demanding in both time and storage. In order to obtain this object-centered representation for an object with n surface points/image pixels, the computation cost will be $O(n^2)$.
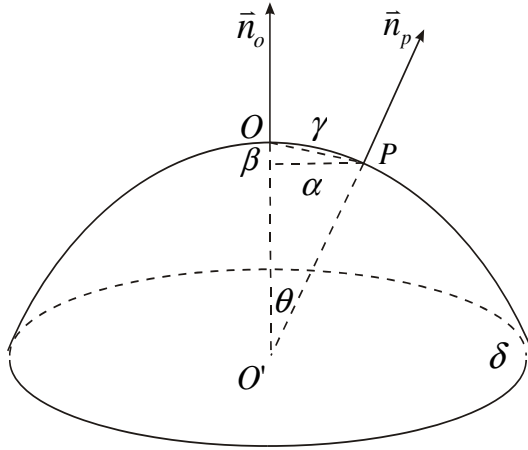
Fig. 1. Illustration of spin image construction

4   Adapting Spin Images to Needle-Maps

We have adapted a patch based approach to spin-image representation. We segment the surface into patches and for each patch we use only the geometric center point O to construct the spin image, rather than use every point of this surface as in the original spin image approach. Our histograms are constructed on a patch-by-patch basis.

From the GGFI [6] we obtain a surface height $\left|\overrightarrow{OP'}\right|$ in the viewing direction $\overrightarrow{n_v}$. The surface height $\left|\overrightarrow{OP'}\right|$ and the horizontal distance $\left|\overrightarrow{P'P}\right|$ can not be used to construct the spin image because they are based on the viewing direction. What we need is the surface height and the horizontal distance based on the surface normal $\overrightarrow{n_o}$ direction.

To compute these quantities we proceed as follows. From the surface height $\left|\overrightarrow{OP'}\right|$, the distance $\left|\overrightarrow{OP}\right|$ is easy to compute since the distance $\left|\overrightarrow{PP'}\right|$ on the viewing plane can be directly measured from the 2D image. From Fig. 2, we are interested in the distance α, i.e. $\left|\overrightarrow{PP''}\right|$, on the surface normal plane δ and the relative height β, i.e. $\left|\overrightarrow{OP''}\right|$, in the surface normal direction $\overrightarrow{n_o}$ between the reference point O and the arbitrary point P. The quantities are related by the following equations:

$$\overrightarrow{OP}=\overrightarrow{OP'} + \overrightarrow{P'P} \tag{3}$$

$$\overrightarrow{P''O}= (\overrightarrow{OP} \cdot \overrightarrow{n_o})\times \overrightarrow{n_o} \tag{4}$$
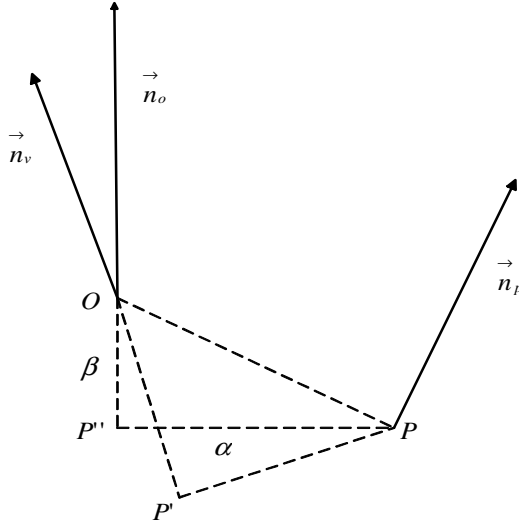
Fig. 2. Illustration of how to obtain the distance α and the relative height β

$$\overrightarrow{P''P} = \overrightarrow{P''O} + \overrightarrow{OP} \tag{5}$$

We now have all the ingredients to construct the 2D histogram of α and β for the surface patch centered at the point O. Equation 6 shows the structure of the patch based spin image, in which M is the number of surface patches and n is the number of histogram bins.

$$
\begin{matrix}
O_1 & O_2 \\
\begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{1n} \\
a_{21} & a_{22} & \ldots & a_{2n} \\
\vdots & & \vdots & \\
a_{n1} & a_{n2} & \ldots & a_{nn}
\end{bmatrix}
&
\begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{1n} \\
a_{21} & a_{22} & \ldots & a_{2n} \\
\vdots & & \vdots & \\
a_{n1} & a_{n2} & \ldots & a_{nn}
\end{bmatrix} \\
\ldots & O_M \\
& \begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{1n} \\
a_{21} & a_{22} & \ldots & a_{2n} \\
\vdots & & \vdots & \\
a_{n1} & a_{n2} & \ldots & a_{nn}
\end{bmatrix} \\
\ldots &
\end{matrix}
\tag{6}
$$

In our experiment we construct a 10 by 10 bin 2D histograms of α and β for an image patch of 32 by 32 pixels. The histogram is also normalised so as to be scale invariant.

As an additional step, we have performed PCA on the spin image histograms to reduce the dimensionality of the data. The idea is as follows. We normalise the contents of each spin image histogram to unity. The normalised bin contents of the histograms are concatenated as a long-vector as follows.

$$
\begin{bmatrix}
\{O_1 : a_{11}\}, \{O_1 : a_{12}\}, \ldots, \{O_1 : a_{1n}\}, \ldots, \\
\{O_1 : a_{n1}\}, \{O_1 : a_{n2}\}, \ldots, \{O_1 : a_{nn}\}, \\
\{O_2 : a_{11}\}, \{O_2 : a_{12}\}, \ldots, \{O_2 : a_{1n}\}, \ldots, \\
\{O_2 : a_{n1}\}, \{O_2 : a_{n2}\}, \ldots, \{O_2 : a_{nn}\}, \\
\ldots, \\
\{O_M : a_{11}\}, \{O_M : a_{12}\}, \ldots, \{O_M : a_{1n}\}, \ldots, \\
\{O_M : a_{n1}\}, \{O_M : a_{n2}\}, \ldots, \{O_M : a_{nn}\}
\end{bmatrix}
\tag{7}
$$

Dimensionality reduction is effected by projecting the long-vector onto the leading eigenvectors of the long-vector covariance matrix.

In the adaptation of spin image on surface normal, the computation cost is reduced to $O(n)$ instead of $O(n^2)$ in the original approach.

5   Recognition

In our preprocessing of the images to extract needle-maps, we perform alignment. This means that we can apply a patch template to the extracted needle-maps to decompose the face into regions. The patch template is constructed from the mean facial needle map, and consists of regions that are either wholly concave or wholly convex. The convexity/concavity test is made using the sign of the changes in surface normal direction. By performing the spin image analysis on these regions, we avoid problems associated with inflexion points when the approximations outlined in Sect. 2 are employed.

As an alternative to constructing the template from the mean needle-map, we have explored constructing it from the needle map extracted from each facial image.

Our measure of facial similarity is based on the normalised correlation of the spin-image histograms for corresponding template patches.

Johnson and Hebert use normalised correlation to evaluate spin image similarity[4]. The method assumes that spin-images from proximal points on the surface for different views of an object will be linearly related. This is because the number of points that fall into corresponding bins will be similar (given that the distribution of points over the surface of the objects is the same). In our case, this assumption still holds. We hence use normalised correlation to compare the patch-based spin images. The correlation is given by

$$
r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}}
\tag{8}
$$

where r_{xy} is the correlation of two spin images x and y. n is the bin number of the spin image, x_i and y_i is the bin contents of two spin images respectively.

$$
r_{sum} = \sum_{i=1}^{min(M,N)} r_{m_i n_i}
\tag{9}
$$

6   Experiments

We apply our method to the CMU PIE face database. We use cropped frontal-viewed face images (without background) in this paper. The sub-database contains $67 \times 7 = 469$ (67 subjects (1-67) and 7 lights (3,10,7,8,9,13,16)) images. We apply the two different patch segmentation strategies outlined above.

For the 7 images of the same subject illuminated by different lights, we use 3 for training sets and 4 for testing. To perform recognition for the 67 subjects, we select a probe image from the test set and the closest image in the training set. The results of our experiments are summarised using the precision-recall curves shown in Fig. 5. The star-dotted curve shows the result of patch-based similarity, the circle-dotted curve shows the result of comparing the vectors extracted using PCA and the cross-dotted curve shows the result of using a global histogram of curvature attributes extracted from the needle-maps [5]. The best results are obtained by applying PCA to the spin image histograms.

In Table 1 we compare the recognition results obtained using the spin-image and applying PCA to the spin image long-vectors. Performance is improved using PCA, and this can be ascribed to the fact that PCA effectively discards the histogram bins that are associated with insignificant variance.

Please notice the face component performance is obtained by only comparing the similarity of a single face component (eye, nose, mouth, etc.) instead of the whole face, so the recognition rate will be reasonably low and can only be used to compare the performance of two methods.

Table 1. Recognition performance using the spin image and the dimension-reduced spin image vector

	Spin Image	Spin Image Vector
Face Components	27.81%	29.69%
Whole Face	87.50%	98.75%

In Table 2 we show the effect of applying the different shape-representations to the initial needle maps and the adjusted needle maps obtained using the mean needle map. In each case there is a significant improvement.

Table 2. Recognition performance using the surface normal aligned to mean needle map and and the original surface normal

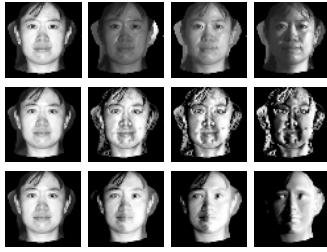|	Original	MNMA
Global Histogram	47.50%	62.50%
Spin Image	85.00%	87.5%
Spin Image Vector	95.00%	98.75%

Fig. 3. The images in the first row are real images illuminated by the light sources from different directions. The images in the second row are the original needle maps rendered by the light source different from the real one [10]. The images in the third row are the needle maps after Mean Needle Map Alignment (Sect. 2) rendered by the light source different from the real one. The images in the third row are more photo-realistic and carry less noise than the ones in the second row.

Fig. 4. The first image is the original needle map projected to the x direction of the viewing plane. The second image is the mean needle map that we use as the template. The third image is the needle map projected to the x direction of the viewing plane after the Mean Needle Map Alignment (Sect. 2). The third image compensate the concave/convex problem of the first one.
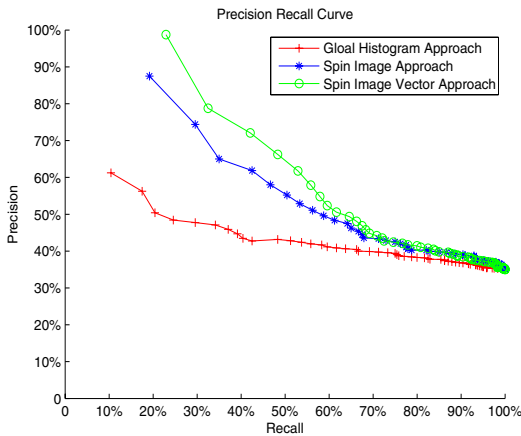

Fig. 5. There are three precision-recall curves of different approaches in this figure: the basic patch-based spin image approach, the spin image vector approach and the previous global histogram approach [5]. All these results are based on the surface normal processed by the Mean Needle Map Alignment because that approach has been proved improving the distinguishing ability in Table 2. Among them the spin image vector approach gives the best performance.

7 Conclusion and Future Work

In this paper we have explored how spin images can can extracted from 2D facial images using shape-from-shading. We make a number of contributions. First, we show how the spin-image histograms can be computed from needle-maps. Second, we show how the complexity of the spin image computation can be reduced using patches, and how the dimensionality of the histograms can be reduced using PCA. Third, we show how the problems of concave/convex inversions in the needle map can be overcome using a mean facial needle-map. The performance of the method is encouraging and can give recognition rates as high as 98.75%.

References

1. P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 28(3):245–260, 1998.
2. Blanz. Automatic face identification system using flexible appearance models. *IVC*, 13(5):393–401, June 1995.
3. V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.
4. A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433–449, 1999.
5. Y. Li and E. Hancock. Face recognition using shading-based curvature attributes.
6. R.T.Frankot and Z.Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions in Pattern Recognition and Machine Intelligence*, 10:439–451, 1988.
7. T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. In *Proceedings of the CVPR 2001 Workshop on Models versus Exemplars in Computer Vision*, December 2001.
8. W. Smith and E. R. Hancock. Recovering facial shape and albedo using a statistical model of surface normal direction. In *Proc. ICCV*, pages 588–595, 2005.
9. M. Turk and A. Pentland. Face recognition using eigenfaces. 1991.
10. P. L. Worthington and E. R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shaping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1250–1267, 1999.