Face Recognition Using a Surface Normal Model

W.A.P. Smith and E.R. Hancock

Department of Computer Science, The University of York
{wsmith, erh}@cs.york.ac.uk

Abstract. This paper describes how facial shape can be modelled using a statistical model that captures variations in surface normal direction. We fit the model to intensity data using constraints on the surface normal direction provided by Lambert's law. We demonstrate that this process yields improved facial shape recovery and can be used for the purposes of illumination insensitive face recognition.

1 Introduction

Although alluring, the use of shape-from-shading for 3D reconstruction and recognition has proved to be an elusive task. This can mainly be attributed to the local convexity-concavity instability due to the bas-relief ambiguity. One way of overcoming this problem with single view shape-from-shading is to use domain specific constraints. Several authors [1,2] have shown that, at the expense of generality, the accuracy of recovered shape information can be greatly enhanced by restricting a shape-from-shading algorithm to a particular class of objects. Zhao and Chellappa [2] introduced a geometric constraint which exploited the approximate bilateral symmetry of faces. Atick et al. [1] proposed a statistical shape-from-shading framework based on a low dimensional parameterisation of facial surfaces.

However, it is surface orientation and not depth which is conveyed by image intensity. It is for this reason that in this paper we investigate whether surface orientation rather than depth can be used to construct a statistical model of face shape. Unfortunately, the construction of a statistical model for the distribution of facial needle-maps is not a straightforward task. The reason for this is that the statistical representation of directional data has proved to be considerably more difficult than that for Cartesian data. For instance, if we consider a short walk across one of the poles of the unit sphere, then although the distance traversed is small, the change in azimuth angle is large.

To overcome the problem, in this paper we draw on ideas from cartography. Our starting point is the *azimuthal equidistant* projection [3]. This projection has the important property that it preserves the distances between between the centre of projection and all other locations on the sphere. Another useful property of this projection is that straight lines on the projected plane through the centre of projection correspond to great circles on the sphere. We exploit these properties to generate a local representation of the field of surface normals. We commence with a set of needle-maps, i.e. fields of surface normals which

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 423–430, 2005.

in practice are obtained either from range images or shape-from-shading. We begin by computing the mean field of surface normals. The surface normals are represented using elevation and azimuth angles on a unit sphere. At each image location the mean-surface normal defines a reference direction which we use to construct an azimuthal equidistant projection for the distribution of surface normals at this point. The distribution of points on the projection plane preserves the distances of the surfaces normals on the unit sphere with respect to the mean surface normal, or reference direction. We then construct a deformable model over the set of surface normals by applying the Cootes and Taylor [4] point distribution model to the co-ordinates that result from transforming the surface normals from the unit sphere to the tangent plane under azimuthal equidistant projection.

The model provides a global statistical constraint which we exploit to help resolve the ambiguity in the shape-from-shading process. In addition, the model parameters describing a recovered needle-map are invariant to changes in illumination. We therefore use these parameters to perform illumination insensitive face recognition.

2 A Statistical Surface Normal Model

A "needle map" describes a surface $z(x,y)$ as a set of local surface normals $\mathbf{n}(x,y)$ projected onto the view plane. Let $\mathbf{n}_k(i,j) = (n_k^x(i,j), n_k^y(i,j), n_k^z(i,j))^T$ be the unit surface normal at the pixel indexed (i,j) in the k^{th} training image. If there are T images in the training set, then at the location (i,j) the mean-surface normal direction is $\hat{\mathbf{n}}(i,j) = \frac{\bar{\mathbf{n}}(i,j)}{||\bar{\mathbf{n}}(i,j)||}$ where $\bar{\mathbf{n}}(i,j) = \frac{1}{T}\sum_{k=1}^{T}\mathbf{n}_k(i,j)$.

On the unit sphere, the surface normal $\mathbf{n}_k(i,j)$ has elevation angle $\theta_k(i,j) = \frac{\pi}{2} - \arcsin n_k^z(i,j)$ and azimuth angle $\phi_k(i,j) = \arctan\frac{n_k^y(i,j)}{n_k^x(i,j)}$, while the mean surface normal at the location (i,j) has elevation angles $\hat{\theta}(i,j) = \frac{\pi}{2} - \arcsin\hat{n}^z(i,j)$ and azimuth angle $\hat{\phi}(i,j) = \arctan\frac{\hat{n}^y(i,j)}{\hat{n}^x(i,j)}$.

To construct the azimuthal equidistant projection we commence by constructing the tangent plane to the unit-sphere at the location corresponding to the mean-surface normal. We establish a local co-ordinate system on this tangent plane. The origin is at the point of contact between the tangent plane and the unit sphere. The x-axis is aligned parallel to the local circle of latitude on the unit-sphere. Under the azimuthal equidistant projection at the location (i,j), the surface normal $\mathbf{n}_k(i,j)$ maps to the point with co-ordinate vector $\mathbf{v}_k(i,j) = (x_k(i,j), y_k(i,j))^T$. The transformation equations between the unit-sphere and the tangent-plane co-ordinate systems are

$$x_k(i,j) = k' \cos\theta_k(i,j)\sin[\phi_k(i,j) - \hat{\phi}(i,j)]$$

$$y_k(i,j) = k'\left\{\cos\hat{\theta}(i,j)\sin\phi_k(i,j) - \sin\hat{\theta}(i,j)\cos\theta_k(i,j)\cos[\phi_k(i,j) - \hat{\phi}(i,j)]\right\}$$

where $\cos c = \sin\hat{\theta}(i,j)\sin\theta_k(i,j) + \cos\hat{\theta}(i,j)\cos\theta_k(i,j)\cos[\phi_k(i,j) - \hat{\phi}(i,j)]$ and $k' = \frac{c}{\sin c}$.

The equations for the inverse transformation from the tangent plane to the unit-sphere are

$$\theta_k(i,j) = \sin^{-1}\left\{\cos c \sin\hat{\theta}(i,j) - \frac{1}{c}y_k(i,j)\sin c\cos\hat{\theta}(i,j)\right\}$$
$$\phi_k(i,j) = \hat{\phi}(i,j) + \tan^{-1}\psi(i,j)$$

where

$$\psi(i,j) = \begin{cases} \frac{x_k(i,j)\sin c}{c\cos\hat{\theta}(i,j)\cos c - y_k(i,j)\sin\hat{\theta}(i,j)\sin c} & \text{if } \hat{\theta}(i,j) \neq \pm\frac{\pi}{2} \\ -\frac{x_k(i,j)}{y_k(i,j)} & \text{if } \hat{\theta}(i,j) = \frac{\pi}{2} \\ \frac{x_k(i,j)}{y_k(i,j)} & \text{if } \hat{\theta}(i,j) = -\frac{\pi}{2} \end{cases}$$

and $c = \sqrt{x_k(i,j)^2 + y_k(i,j)^2}$.

For each image location the transformed surface normals from the T different training images are concatenated and stacked to form two long-vectors of length T. For the pixel location indexed (i,j), the first of these is the long vector with the transformed x-co-ordinates from the T training images as components, i.e. $\mathbf{V}_x(i,j) = (x_1(i,j), x_2(i,j), ..., x_T(i,j))^T$ and the second long-vector has the y coordinate as its components, i.e. $\mathbf{V}_y(i,j) = (y_1(i,j), y_2(i,j), ..., y_T(i,j))^T$. Since the equidistant azimuthal projection involves centering the local co-ordinate system, the mean long-vectors over the training images are zero. If the data is of dimensions M rows and N columns, then there are $M \times N$ pairs of such long-vectors. The long vectors are ordered according to the raster scan (left-to-right and top-to-bottom) and are used as the columns of the $T \times (2MN)$ data-matrix $\mathbf{D} = (\mathbf{V}_x(1,1)|\mathbf{V}_y(1,1)|\ \mathbf{V}_x(1,2)|\mathbf{V}_y(1,2)|\ \ldots|\mathbf{V}_x(M,N)|\mathbf{V}_y(M,N))$. The covariance matrix for the long-vectors is the $(2MN)\times(2MN)$ matrix $\mathbf{L} = \frac{1}{T}\mathbf{D}^T\mathbf{D}$. We follow Atick et al. [1] and use the numerically efficient method of Sirovich [5] to compute the eigenvectors $\mathbf{e}_i$ of $\mathbf{L}$. We deform the equidistant azimuthal point projections in the directions defined by the $2MN\times K$ matrix $\mathbf{P} = (\mathbf{e}_1|\mathbf{e}_2|\ldots|\mathbf{e}_K)$ formed from the leading K principal eigenvectors. This deformation displaces the transformed surface normals on the local tangent planes in the directions defined by the eigenvectors $\mathbf{P}$. If $\mathbf{b} = (b_1, b_2, ..., b_K)^T$ is a vector of parameters of length K, then since the mean-vector of co-ordinates resulting from the equidistant azimuthal projection is zero, the deformed vector of projected co-ordinates is $\mathbf{v}_b = \mathbf{P}\mathbf{b}$. Suppose that $\mathbf{v}_o$ is the vector of co-ordinates obtained by performing the azimuthal equidistant projection on an observed field of surface normals. We seek the parameter vector $\mathbf{b}$ that minimises the squared error $\mathcal{E}(\mathbf{b}) = (\mathbf{v}_o - \mathbf{P}^T\mathbf{b})^T(\mathbf{v}_o - \mathbf{P}^T\mathbf{b})$. The solution to this least-squares estimation problem is $\mathbf{b}^* = \mathbf{P}^T\mathbf{v}_o$. The best fit field of surface normals allowed by the model is $\mathbf{v}_o^* = \mathbf{P}\mathbf{P}^T\mathbf{v}_o$. The deformed vector of azimuthal equidistant projection co-ordinates can be transformed back into a surface normal on the unit sphere using the inverse azimuthal equidistant projection equations given above.

3 Fitting the Model to Intensity Images

We may exploit the statistical constraint provided by the model in the process of fitting the model to an intensity image and thus help resolve the ambiguity in the shape-from-shading process. We do this using an iterative approach which can be posed as that of recovering the best-fit field of surface normals from the statistical model, subject to constraints provided by the image irradiance equation.

If I is the measured image brightness, then according to Lambert's law $I = \mathbf{n}.\mathbf{s}$, where $\mathbf{s}$ is the light source direction. In general, the surface normal $\mathbf{n}$ can not be recovered from a single brightness measurement since it has two degrees of freedom corresponding to the elevation and azimuth angles on the unit sphere. In the Worthington and Hancock [6] iterative shape-from-shading framework, data-closeness is ensured by constraining the recovered surface normal to lie on the reflectance cone whose axis is aligned with the light-source vector $\mathbf{s}$ and whose opening angle is $\alpha = \arccos I$. At each iteration the surface normal is free to move to an off-cone position subject to smoothness or curvature consistency constraints. However, the hard irradiance constraint is re-imposed by rotating each surface normal back to its closest on-cone position. This process ensures that the recovered field of surface normals satisfies the image irradiance equation after every iteration.

Suppose that $\mathbf{n'}^l(i,j)$ is an off-cone surface normal at iteration l of the algorithm. The update equation is therefore $\mathbf{n}^{l+1}(i,j) = \Theta\mathbf{n'}^l(i,j)$ where Θ is a rotation matrix computed from the apex angle α and the angle between $\mathbf{n'}^l(i,j)$ and the light source direction $\mathbf{s}$. To restore the surface normal to the closest on-cone position it must be rotated by an angle $\theta = \alpha - \arccos\left[\mathbf{n'}^l(i,j).\mathbf{s}\right]$ about the axis $(u,v,w)^T = \mathbf{n'}^l(i,j) \times \mathbf{s}$. Hence, the rotation matrix is

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

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

The framework is initialised by placing the surface normals on their reflectance cones such that they are aligned in the direction opposite to that of the local image gradient (biasing towards global convexity).

Our approach to fitting the model to intensity images uses the fields of surface normals estimated using the geometric shape-from-shading method described above. This is an iterative process in which we interleave the process of fitting the statistical model to the current field of estimated surface normals, and then re-enforcing the data-closeness constraint provided by Lambert's law by mapping the surface normals back onto their reflectance cones. The algorithm can be summarised as follows:

1. Calculate an initial estimate of the field of surface normals $\mathbf{n}$ by aligning each normal on its reflectance cone with the negative local intensity gradient.

2. Each normal in the estimated field $\mathbf{n}$ undergoes an azimuthal equidistant projection to give a vector of transformed coordinates $\mathbf{v}_o$.
3. The vector of best fit model parameters is given by $\mathbf{b} = \mathbf{P}^T\mathbf{v}_o$.
4. The vector of transformed coordinates corresponding to the best-fit parameters is given by $\mathbf{v}' = \mathbf{P}\mathbf{P}^T\mathbf{v}_o$.
5. Using the inverse azimuthal equidistant projection find the off-cone best fit surface normal $\mathbf{n}'$ from $\mathbf{v}'$.
6. Find the on-cone surface normal $\mathbf{n}''$ by rotating the off-cone surface normal $\mathbf{n}'$ using $\mathbf{n}'' = \mathbf{\Theta}\mathbf{n}'$.
7. Test for convergence. If $\sum_{i,j} \cos^{-1}\left[\mathbf{n}(i,j).\mathbf{n}''(i,j)\right] < \epsilon$, where ϵ is a predetermined threshold, then stop and return $\mathbf{b}$ as the estimated model parameters and $\mathbf{n}''$ as the recovered needle map.
8. Make $\mathbf{n} = \mathbf{n}''$ and return to step 2.

Since real world face images contain albedo variations, we choose to output $\mathbf{n}'$ and estimate the facial albedo map using the differences between observed and reconstructed image brightness, i.e. we relax the data-closeness constraint at the final iteration. Hence, the albedo ρ is given by $\rho(i,j) = \frac{I(i,j)}{\mathbf{s}.\mathbf{n}'(i,j)}$, where $\mathbf{s}$ is the light source vector.

4 Experiments

In Figure 1 we illustrate the results of the model fitting process. We train the statistical model using surface normals extracted from 200 range images of male and female subjects in frontal poses and neutral expressions [7]. We fit the statistical model to an image using the technique described in Section 3. As input we use images of 10 subjects from the Yale B database [8] in frontal pose and illuminated by a single light source with direction $[0\ 0\ 1]^T$. The algorithm typically converged within 20 iterations. We show the surfaces recovered by integrating the best fit needle maps using the technique of Frankot and Chellappa [9]. In the first and third rows the surfaces are shown rotated 30° about the vertical axis. The surfaces are rendered with Lambertian reflectance and the estimated albedo maps. The light source remains fronto-parallel with respect to the face. The resulting synthesised images are near photo-realistic under a large change in viewpoint. Certainly, the results are comparable with those of Georghiades et al. [8] in which 7 input images were required per subject. The second and fourth rows of Figure 1 show the meshes of the recovered surfaces to allow inspection of the recovered shape alone. In Figure 2 we demonstrate that the recovered surface and albedo map are sufficiently stable to synthesise images in both novel pose and novel illumination. We show the surface of the eighth subject from the previous figure and circle the light source from left profile to right profile.

Provided that the shape-from-shading process is sufficiently accurate, the parameters describing a recovered facial needle map are invariant to illumination and reflectance properties. They hence encode only appearance. For this reason these parameters potentially provide a means of performing illumination and reflectance invariant face recognition. In our experiments we use a

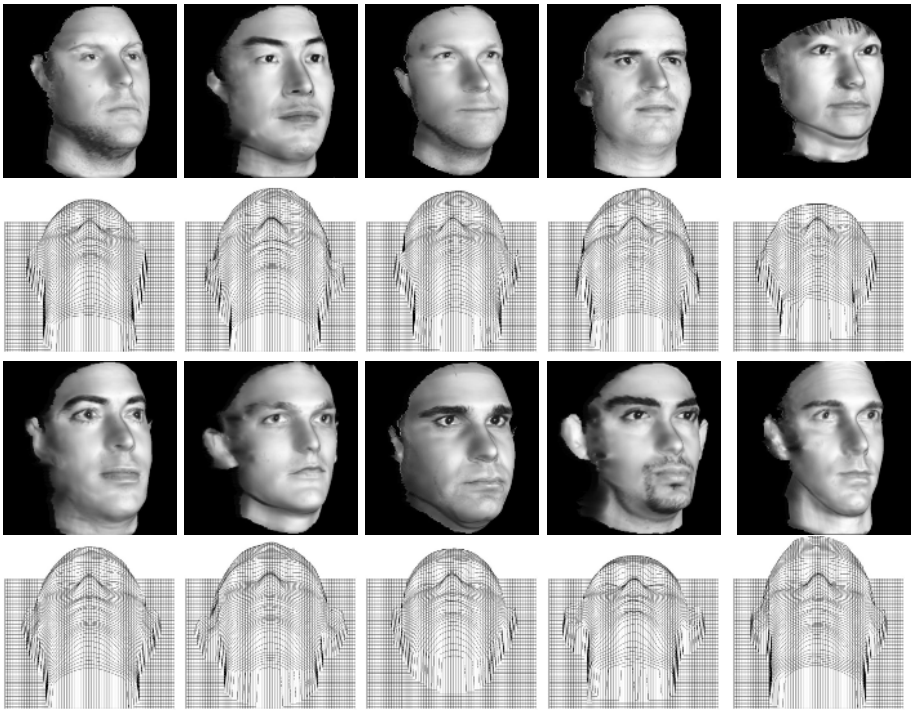


Fig. 1. Surfaces recovered from the ten subjects in the Yale B database

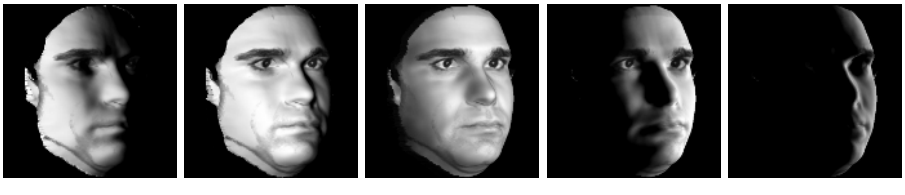


Fig. 2. Surface recovered from subject 8 in novel pose and under varying illumination

subset of the CMU PIE database [10]. This database contains images of 67 subjects under varying pose, illumination and expression. We fix the pose to fronto-parallel (camera c27) and vary the illumination direction along a horizontal arc approximately 55° in each direction. The set of flashes used was {f03,f10,f07,f08,f09,f13,f16}. For each subject we use only one gallery image, illuminated from close to the viewing direction (flash f08). We fit the statistical model to each gallery image which provides an appearance vector for each subject. For each probe image we repeat the same process to find an appearance vector for the unknown face.

In order to effect recognition with 1 gallery image per subject, we find the Euclidian distances between a given probe vector and all the gallery vectors and sort them. A probe has rank k if the correct match is the k^{th} smallest Euclidian

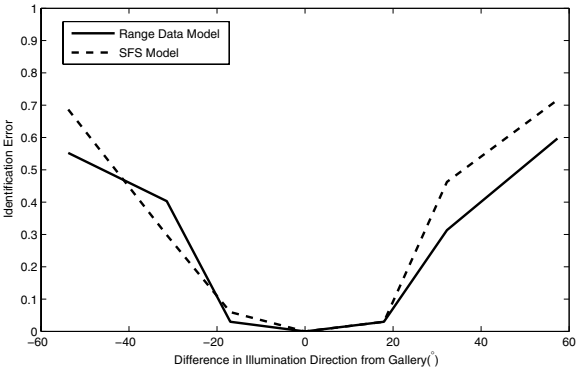
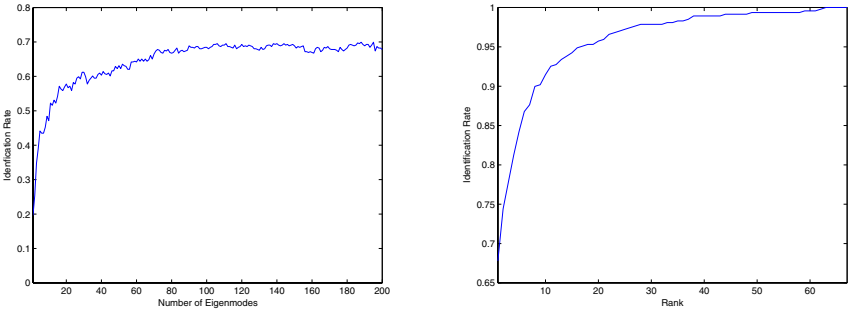


Fig. 3. Recognition error versus angle of illumination



(a) (b)

Fig. 4. Identification rate versus (a) the number of Eigenmodes used and (b) rank

distance. A probe is correctly identified if it has rank 1, i.e. the correct match has the lowest Euclidian distance. The rank can vary between 1 and the size of the gallery (i.e. 67). In Figure 3 we begin by showing the rank 1 identification error as the illumination angle is varied through approximately 110°. We experiment with both the model trained on ground truth normals extracted from range data and on normals extracted using shape-from-shading.

From the plot it is clear that low error rates are achievable for variations in illumination direction of approximately ±30°. Beyond this point, performance decreases rapidly. We believe this is because as the illumination direction becomes more extreme, increasing areas of the face are in shadow. Thus, the imposition of data-closeness at each iteration in these shadow areas may result in the fitting process 'walking away' from the true solution. Under these conditions, the iterative fitting process would be best placed in a statistical setting which attempts to match the model to the needle map in the visible areas only.

In Figure 4 (a) we investigate how the number of eigenmodes used affects identification performance. We show the total rank 1 identification rate across all illumination conditions against the number of eigenmodes used. Performance

appears to level out at around 100 eigenmodes (dimensions), suggesting that additional dimensions do not encode modes of facial shape which are useful for recognition.

The identification rate at rank k is the fraction of probes that have rank k or lower. Plotting identification performance against rank on a cumulative match characteristic (CMC) allows us to asses whether the correct match is in the top k matches rather than just the first. Most applications are based on performance at lower ranks. From the CMC in Figure 4 (b) it can be seen that performance rises sharply from rank 1 to 10.

5 Conclusions

We have shown how a statistical model of shape may be constructed from fields of surface normals using the azimuthal equidistant projection. We presented an iterative method for fitting the model to an image subject to image irradiance constraints. The method proves rapid to converge, and delivers realistic surfaces when the fields of surface normals are integrated. The resulting parameter vector provides a means to perform illumination insensitive face recognition. The technique could also be use as a generative model to sample the entire pose and illumination space from a single image. Our future plans revolve around placing the iterative process in a statistical setting in order to improve the reliability of the estimated parameters under varying illumination. We also plan to develop ways of aligning the model with images which are not in a frontal pose.

References

1. Atick, J.J., Griffin, P.A., Redlich, A.N.: Statistical approach to SFS: Reconstruction of 3D face surfaces from single 2D images. Neural Comp. **8** (1996) 1321–1340
2. Zhao, W.Y., Chellappa, R.: Illumination-insensitive face recognition using symmetric SFS. In: Proc. CVPR. (2000)
3. Snyder, J.P.: Map Projections–A Working Manual, U.S.G.S. Professional Paper 1395. United States Government Printing Office, Washington D.C. (1987)
4. Cootes, T.F., J.Taylor, C., Cooper, D., Graham, J.: Training models of shape from sets of examples. In: Proc. BMVC. (1992) 9–18
5. Sirovich, L.: Turbulence and the dynamics of coherent structures. Quart. Applied Mathematics **XLV** (1987) 561–590
6. Worthington, P.L., Hancock, E.R.: New constraints on data-closeness and needle map consistency for shape-from-shading. IEEE Trans. PAMI **21** (1999) 1250–1267
7. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Computer Graphics Proc. SIGGRAPH. (1999) 187–194
8. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. PAMI **23** (2001) 643–660
9. Frankot, R.T., Chellappa, R.: A method for enforcing integrability in shape from shading algorithms. IEEE Trans. PAMI **10** (1988) 439–451
10. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression database. IEEE Trans. Pattern Analysis and Machine Intelligence **25** (2003) 1615–1618