



ELSEVIER

Pattern Recognition Letters 23 (2002) 533–548

---

---

Pattern Recognition  
Letters

---

---

www.elsevier.com/locate/patrec

# Estimating facial pose using shape-from-shading

Kwang Nam Choi, Philip L. Worthington, Edwin R. Hancock \*

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

Received 25 August 2000; received in revised form 24 April 2001

---

## Abstract

This paper reports the application of a recently developed shape-from-shading technique to estimate facial pose. The shape-from-shading algorithm uses a new geometric technique for solving the image irradiance equation together with curvature consistency constraints. Orientation histograms extracted from the the needle-maps delivered by the new shape-from-shading algorithm are used to estimate facial pose. We present a simple model of how the histogram bin-contents transform under rotation of the head. The parameters of this model are the head pose angles. We estimate pose by searching for the rotation angles which maximise the correlation between transformed histograms. A sensitivity analysis reveals that the methods can deliver pose estimates that are accurate within a few degrees. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Face recognition; Pose estimation; Shape-from-shading

---

## 1. Introduction

The use of shading and illumination information for face analysis has been a longstanding goal in computational vision. The topic has attracted considerable interest in both the psychology and computer vision literature. In the psychology literature, several authors have investigated the effects of orientation, albedo and shading and have established their importance for face perception and recognition. For instance, Enns and Shore (1997) have studied the influence of both orientation and lighting on the inverted-face effect. Kemp

et al. (1996) have studied the effects of shading and pigmentation on the perception and recognition of faces. Braje et al. (1998) have studied the role of illumination effects on face recognition. In the computer vision area, Troje and Bulthoff (1996) investigated the role of facial shape and texture in the recognition of shaded faces. Yuille et al. (1998) have shown how shading information can be used to control the warping of a prototype shape onto image data. Atick et al. (1996) show how principal component analysis (PCA) can be used to reconstruct 3D facial surfaces from 2D images.

This paper addresses the problem of whether shading information can be used to facial pose estimation. Facial pose estimation is a topic of practical importance, since the estimation of gaze direction and facial attitude is of crucial importance in the design of automatic surveillance

---

\* Corresponding author. Tel.: +44-1904-432722; fax: +44-1904-432767.

*E-mail address:* erh@cs.york.ac.uk (E.R. Hancock).

systems and vision-based interfaces. There is a considerable body of work on the topic. Most of this is based either on the use of the locations of facial features or on template based methods, rather than the use of information provided by shading patterns. For instance, Gee and Cipolla (1994) estimated the direction of gaze from a single view of a face using the positions of the eyes. The method requires very few image measurements and can be run in real-time. Hattori et al. (1998) have achieved facial pose estimation using a much richer set of facial features derived from colour and range images. Wiskott et al. (1996) have shown how the bunch graph representation can be used to recover pose information using a relational model. Several authors have couched the pose estimation problem as one of learning from examples. McKenna et al. (1996) have used PCA to examine the distribution of face pose without using facial features. Huang et al. (1997, 1998) have used support vector machines (SVMs). Romdhani et al. (1999) have proposed a multi-view nonlinear active shape model (Cootes et al., 1994) that uses 2D view-dependent contextual constraints without explicit reference to 3D object geometry of the head of a human figure.

The aim in this paper is to investigate whether shape-from-shading can be used for the purposes of estimating facial pose. Because of its perceived fragility, the use of surface orientation or needle-map information has received little interest in the literature. One of the reasons for this is that shape-from-shading has proved notoriously difficult to regulate and the recovered surface orientation information is of questionable acuity. In particular, the recovered needle-maps are dominated by the smoothness constraints needed to solve the image irradiance equation. As a result, the image irradiance equation is only weakly satisfied and the recovered orientation information is a poor representation of reality. For this reason, much of the face-analysis literature has focused on recovering shape from multiple light source images using photometric stereo techniques or by developing more sophisticated appearance based models to account for the observed shading variations. For example, Belhumeur and Kriegman (1996) have developed a photometric stereo method which relies on matrix factorisation to

extract shape and albedo information from multiple light source images of the human face. Jacobs et al. (1998) have shown how to use this representation to compare faces under different illumination conditions. More recently, Belhumeur et al. (1997) have shown how to use a class specific linear projection to recognise different faces under variations in illumination. Georghiades et al. (1998) have a more physically based method which allows the illumination cone to be used to effect recognition under variable lighting conditions when a Lambertian reflectance model applies.

While this work has certainly proved effective, we make the observation that it is based on using multiple light source information and is not concerned with using shading information from a single image of a face. Worthington and Hancock (1999) have recently reported a new framework for shape-from-shading which holds out the possibility of performing shaded face analysis using a single image. Their main contribution has been to develop an efficient geometric means of iteratively adjusting the needle-map orientation, which ensures that the image irradiance equation can be satisfied as a hard constraint. Moreover, they have developed alternatives to the usual assumption of quadratic smoothness which allow gradient consistency and curvature consistency constraints to be imposed on the recovered needle-maps. The net effect is to recover orientation information, which is a more faithful representation of ground-truth.

The aim in this paper is to investigate whether the local surface orientation information delivered by the new shape-from-shading technique can be used for facial analysis. Although such information could potentially be used for a number of tasks including recognition and facial feature extraction, in this paper we concentrate on the problem of recovering facial pose. The idea underpinning our facial pose estimation method is a simple one. We aim to model the way in which the distribution of needle-map directions transforms under rotation of the head. By building a simple geometric model, we calculate how the histogram of surface normal slant and tilt angles changes under horizontal and vertical rotations of the head. The pose is recovered by locating the rotation angles that lead to maximum similarity be-

tween the observed orientation histogram and a transformed reference histogram. The reference histogram is obtained from either a face viewed in the fronto-parallel pose, or from a simple model of how the surface normals are distributed when the face is in the fronto-parallel pose. We investigate the systematics of the histogram alignment process. The main conclusion is that the best pose estimates are obtained when the histogram similarity measure is the Bhattacharyya coefficient (Devijver and Kittler, 1982) and a uniform initial distribution of orientations is assumed.

The outline of this paper is as follows. In Section 2 we present the shape-from-shading approach. In Section 3 we describe the orientation histograms, surface normal sample selection, pose model and a method to estimate pose parameters. Experimental evaluation of the method is presented in Section 4. The process of locating rotation angles is also presented in Section 4. Finally, conclusions are drawn in Section 5.

## 2. Shape-from-shading

For completeness, in this section we detail the shape-from-shading algorithm of Worthington and Hancock (1999) which we have used in our experiments. This algorithm has been demonstrated to deliver needle-maps which preserve fine surface detail. 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 neigh-

bours. 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  $s$  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

$$E_{i,j} = \mathbf{n}_{i,j}^k \cdot \mathbf{s}. \quad (1)$$

As a result, the reflectance cone has opening angle  $\cos^{-1} E(i, j)$ . After local smoothing, the off-cone surface normal is  $\bar{\mathbf{n}}_{i,j}^k$ . The updated on-cone surface normal which satisfies the image irradiance equation as a hard constraint is obtained via the rotation

$$\mathbf{n}_{i,j}^{k+1} = \Phi \bar{\mathbf{n}}_{i,j}^k. \quad (2)$$

The matrix  $\Phi$  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

$$\theta = \cos^{-1} E - \cos^{-1} \frac{\bar{\mathbf{n}}_{i,j}^k \cdot \mathbf{s}}{\|\bar{\mathbf{n}}_{i,j}^k\| \cdot \|\mathbf{s}\|}. \quad (3)$$

This rotation takes place about the axis whose direction is given by the vector

$$(u, v, w)^T = \bar{\mathbf{n}}_{i,j}^k \times \mathbf{s}. \quad (4)$$

This rotation axis is perpendicular to both the light source direction and the off-cone normal. Hence, the rotation matrix is

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

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

The off-cone surface normal is recovered through a process of robust-smoothing. The smoothness error or consistency of the field of surface normals is measured using the derivatives of the needle-map in the  $x$  and  $y$  directions by the penalty function

$$I = \int \int \left\{ \left( \rho_\sigma \left( \left\| \frac{\partial \mathbf{n}}{\partial x} \right\| \right) + \rho_\sigma \left( \left\| \frac{\partial \mathbf{n}}{\partial y} \right\| \right) \right) \right\} dx dy. \quad (5)$$

In the above measure,  $\rho_\sigma(\eta)$  is the robust-error kernel used to gauge the local consistency of the needle-map or field of surface normals. The argument of the kernel  $\eta$  is the measured error and the parameter  $\sigma$  controls the width of the kernel. It is important to note that the robust-error kernels are applied separately to the magnitudes of the derivatives of the needle-map in the  $x$  and  $y$  directions. Applying variational calculus, the penalty function is minimised by updating the off-cone surface normals using the following iteration equation

$$\begin{aligned} \bar{\mathbf{n}}_{i,j}^{(k+1)} = & \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial x} \right\|^{-1} \left[ \frac{\partial}{\partial x} \left( \rho'_\sigma \left( \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial x} \right\| \right) \right) \right. \\ & + \rho'_\sigma \left( \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial x} \right\| \right) \times \left( \mathbf{n}_{i+1,j}^{(k)} + \mathbf{n}_{i-1,j}^{(k)} \right) \\ & \left. - \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial x} \right\|^{-2} \left( \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial x} \cdot \mathbf{n}_{i,j}^{(k)} \right) \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial x} \right] \\ & + \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial y} \right\|^{-1} \left[ \frac{\partial}{\partial y} \left( \rho'_\sigma \left( \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial y} \right\| \right) \right) \right. \\ & + \rho'_\sigma \left( \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial y} \right\| \right) \times \left( \mathbf{n}_{i,j+1}^{(k)} + \mathbf{n}_{i,j-1}^{(k)} \right) \\ & \left. - \left\| \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial y} \right\|^{-2} \left( \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial y} \cdot \mathbf{n}_{i,j}^{(k)} \right) \frac{\partial \mathbf{n}_{i,j}^{(k)}}{\partial y} \right]. \quad (6) \end{aligned}$$

Stated in this way the smoothing process is entirely general. However, the most effective error kernel has been found to be the log cosh sigmoidal-derivative M-estimator

$$\rho_\sigma(\eta) = \frac{\sigma}{\pi} \log \cosh \left( \frac{\pi \eta}{\sigma} \right). \quad (7)$$

### 3. Pose estimation

In this section we describe how the needle-maps delivered by Worthington and Hancock's shape-from-shading scheme, described in the previous section, can be used for facial pose estimation.

#### 3.1. Orientation histograms

Our aim is to recover facial pose using orientation histograms extracted from the needle-maps delivered by shape-from-shading. To this end, we compute the angle between the recovered local surface normal  $\mathbf{n} = (p, q, 1)^T$  and the normal-vector  $\mathbf{z} = (0, 0, 1)^T$  to the image plane.

We measure the orientation of the needle-map using two angles. The first of these is the slant angle of the surface normal with respect to the image plane normal. This angle is equal to

$$\alpha = \arctan p. \quad (8)$$

The second angle is the tilt angle with respect to the image plane surface normal

$$\beta = \arctan q. \quad (9)$$

These two angles are binned in a 2D histogram for a sample of surface normals. The normalised contents for the bin-centred on the rotation angles  $\alpha$  and  $\beta$  is denoted by  $h(\alpha, \beta)$ . We have worked with a binning size of one degree in both dimensions.

#### 3.2. Surface normal sample

The sample of surface normals used in our pose estimation algorithm is selected from the forehead. The reasons for choosing this facial region are threefold. Firstly, it is devoid of topographic structure such as ridges and ravines. Secondly, it is approximately planar. Thirdly, it does not suffer occlusion by other facial features. Although other facial regions, e.g., the cheeks, are relatively devoid of structure, we did not use

them for pose estimation since they are subject to occlusion.

The forehead region is a parallelogram, which is selected as follows. The left-most and right-most extremes are vertical lines defined by the centre-points of the eyes. The lower-boundary of the parallelogram is the line that connects the tops of the eyebrows. The upper-boundary is a parallel line that touches the hairline. The feature-points used to define this region are selected manually. The definition of the region of interest and some examples of its automatic segmentation are shown in Fig. 1. In our experiments, we select the region by hand. However, we have recently developed a matched filtering technique which could potentially automatically detect the facial features needed to define the parallelogram.

In Fig. 2 we show the orientation information extracted from the face shown in Fig. 1. The plot shows the slant angle (slant)  $\alpha$  visualised as a grey-scale value. At each pixel the brightness is proportional to the angle  $\alpha$  (slant); the darkest pixels

correspond to large negative values of the angle, while the brightest pixels correspond large positive value. The face used in this example is in the fronto-parallel pose. Notice how the brightness increases from left-to-right across the face. There is a more rapid variation as the nose is traversed. In the panel on the right-hand side of the figure we show the surface normal for the selected forehead region. Notice that this varies relatively smoothly when compared to the remainder of the face.

### 3.3. Pose model

Our pose estimation algorithm is based on correlating the normalised histogram bin-contents with a simple model of the effect of area sampling under angular projection. The facial pose is defined by two angles. The first of these is the rotation in the horizontal direction  $\theta$ . The second angle is the vertical rotation  $\phi$ . As the head rotates, the histogram bin-contents transform. There are two factors controlling the sampling of the needle-map



Fig. 1. Window used for pose estimation.

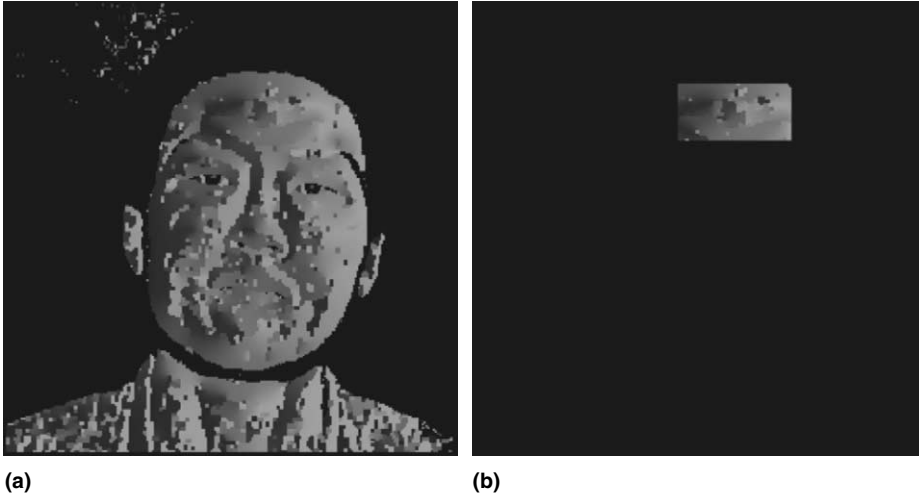


Fig. 2. Surface orientation obtained using SFS: (a) whole face; (b) face window on the forehead for histogram.

histogram. Firstly, the locations of the histogram bin are translated by an amount equal to the angle difference induced by the two components of rotation. Secondly, the contents of the bins are scaled due to the fact that the visible area on the surface of the face is projected onto the image plane. Consider the effect of these two processes on the histogram bin  $h(\alpha, \beta)$ . After rotation of the head the contents will appear in the new bin at location  $(\alpha + \theta, \beta + \phi)$  in the histogram. The effect of area projection results in a scaled bin-contents

$$h_{\theta, \phi}(\alpha + \theta, \beta + \phi) = h_{0,0}(\alpha, \beta) \left| \frac{\sin(\alpha + \theta) \sin(\beta + \phi)}{\sin(\alpha) \sin(\beta)} \right|, \quad (10)$$

where  $h_{0,0}$  is the reference histogram and  $h_{\theta, \phi}$  is the histogram for the rotated head.

### 3.4. Estimating pose parameters

We estimate pose by searching for the rotation angles  $\theta$  and  $\phi$  which maximise the similarity between the measured histogram  $g(\alpha, \beta)$  for a face of unknown pose and the transformed version of a reference histogram for a face of known pose. The reference histogram can be elicited in a number of ways. The simplest method is to

compute the histogram for a fronto-parallel view of the face. This means that we need to construct a reference histogram for each subject studied. Another approach adopted in our experiments is to assume a flat histogram. This is clearly a very poor model of the distribution of orientation angles across an approximately planar surface region. This offers the advantage that we do not need to construct a separate histogram for each subject under study.

To recover pose parameters using the reference histogram, we have investigated two similarity measures. In the first case, we recover pose parameters on the basis of minimum Euclidean distance between histograms, i.e.,

$$\begin{pmatrix} \theta \\ \phi \end{pmatrix} = \arg \min_{\theta, \phi} \sum_{\alpha} \sum_{\beta} [h_{\theta, \phi}(\alpha, \beta) - g(\alpha, \beta)]^2. \quad (11)$$

The second similarity measure is the Bhattacharyya coefficient (Devijver and Kittler, 1982)

$$\begin{pmatrix} \theta \\ \phi \end{pmatrix} = \arg \max_{\theta, \phi} \sum_{\alpha} \sum_{\beta} \sqrt{h_{\theta, \phi}(\alpha, \beta) g(\alpha, \beta)}. \quad (12)$$

In summary, the algorithmic steps of our pose estimation method are:

- Using known light source direction determines the local surface orientation at each point on

the face using the shape-from-shading algorithm outlined in Section 2.

- Isolate a parallelogram on the forehead defined by the feature-points shown in Fig. 2.
- From the sample of surface normals enclosed within this parallelogram compute the 2D histogram of the angles  $\alpha$  and  $\beta$ .
- Search for the set of pose angles which result in either maximum Bhattacharyya coefficient or

minimum Euclidean distance between histograms.

#### 4. Experiments

In this section, we present some experimental evaluation of our new pose estimation method. We commence by describing the data used in our

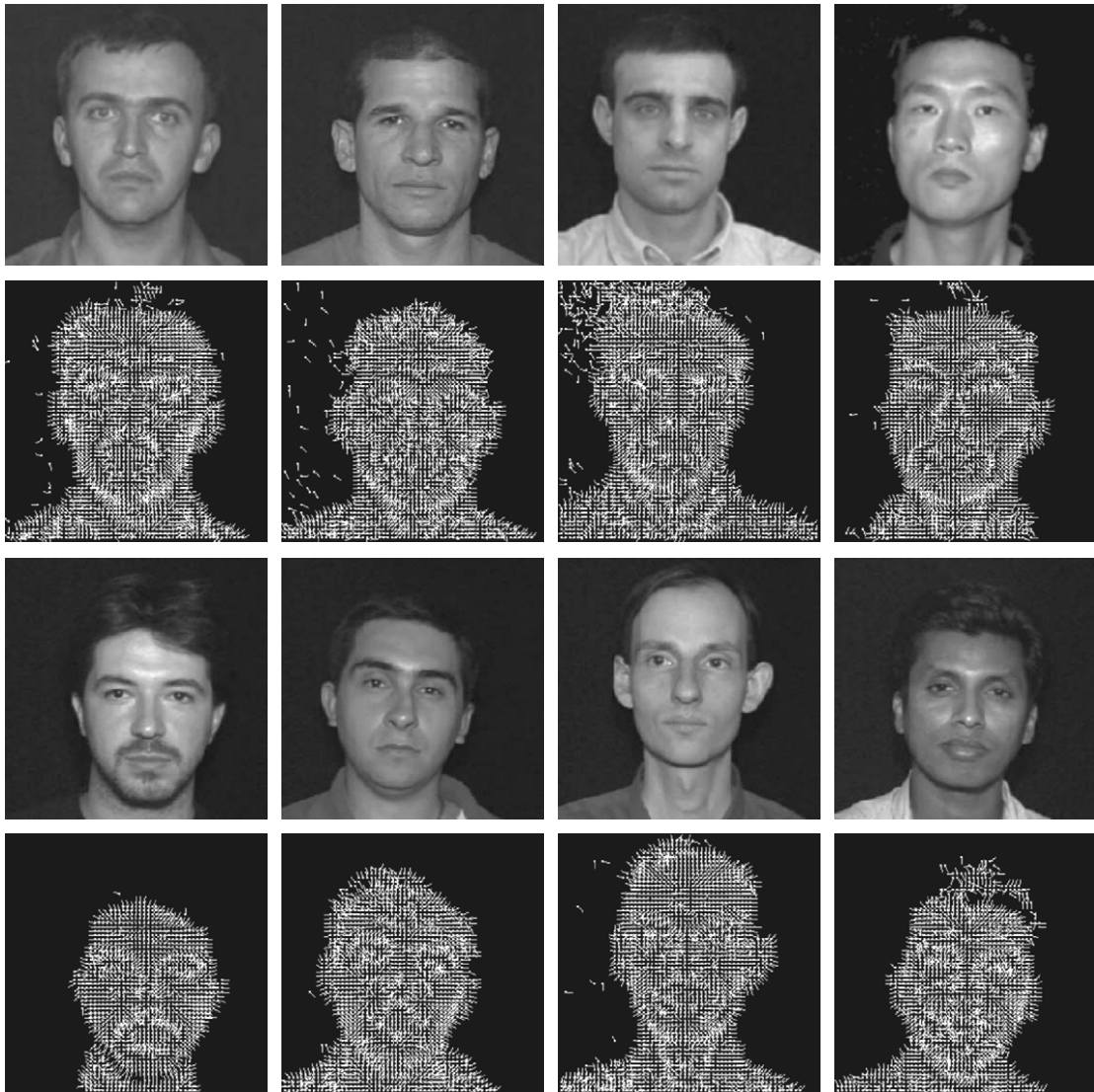


Fig. 3. Example views of natural faces and their needle-maps. The camera direction is approximately fronto-parallel with a single light source.

study. This provides some qualitative examples of the needle-maps used in our study. Second, we provide some evaluation of the accuracy of the new pose estimation method.

#### 4.1. Facial needle-maps

We use two types of data in our experiments. The first of these are images of human subjects under controlled lighting conditions. This data are provided by 21 views of each of eight subjects. The 21 views consist of seven movements of the head from left-to-right (i.e., a clockwise rotation of the head) when the face is inclined upwards, pointing forwards and inclined downwards. It should be noted that several of these images exhibited specularities. We made no attempt to remove these effects at the time the images were captured by, for instance, applying powder to the subjects faces. The second set of data is provided by images of a plaster bust which is rotated by known angles on a turntable between

$-30^\circ$  and  $+30^\circ$ . We use this second source of data to gain ground-truth pose information to measure the sensitivity of our method. The images of the plaster bust were free of specularities.

To control the light source direction used to illuminate the faces, we placed a single fixed light located in front of the face. Here we used an ordinary desk lamp and collimated the light using a cardboard tube. Our reason for doing this was that we did not have access to a point or collimated light source. Better results could probably be obtained using images collected under more carefully controlled lighting conditions. When the images were captured, the light source and camera directions were identical. To avoid problems with multiple light source, the images were all taken in a blacked-out studio. To simplify the problem of localising the faces, the faces were photographed with a black backdrop.

It is worth mentioning at this point that the shape-from-shading method described in this pa-

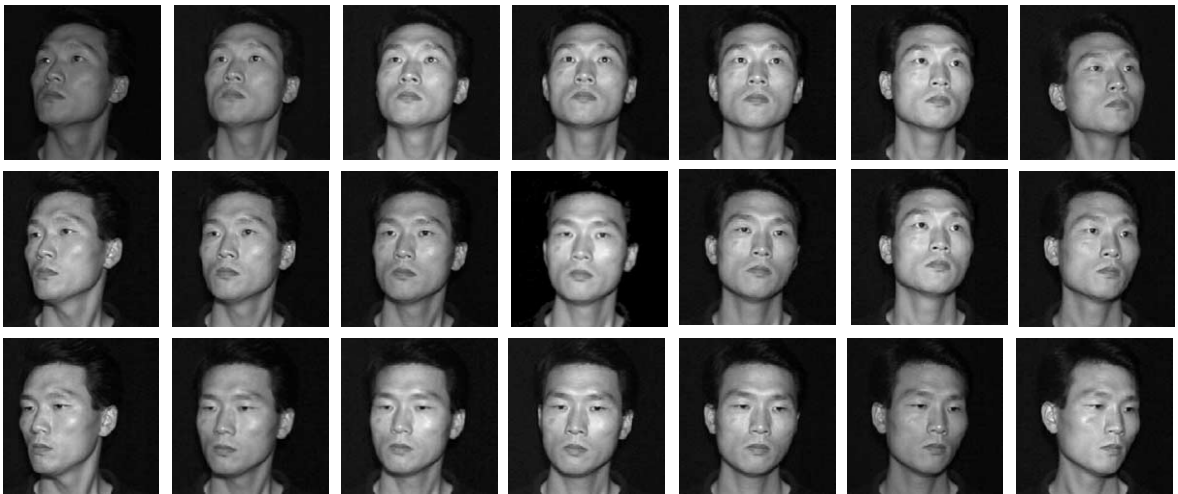


Fig. 4. A series of views of a real-face. The light direction is approximately frontal.



Fig. 5. A series of views of a model-head.

per assumes constant surface albedo. Natural faces clearly depart from this assumption. The lips, eyes and eyebrows are all of different albedo, and some subjects have freckles or other variations in pigmentation.

We now show some examples of the images used in our study. In Fig. 3 we show some of the subjects used in our study together with the recovered needle-maps. The prominent topographic features such as the nose, eye-sockets and lips are well defined. Moreover, the recovered needle-maps are not unduly distorted by the specularities on some of the faces. Fig. 4 shows the sequence of 21 images for one of the subjects used in this study. Fig. 5 is a second sequence for the model head. The corresponding sequences of needle-maps are shown in Figs. 6 and 7. In both cases the topographic features such as the nose and eye-sockets are stable under varying pose.

Next we present some experimental data aimed at establishing the usefulness of the orientation information contained within the needle-maps for pose estimation. Fig. 8 shows the surface normal distributions for three different facial regions. The first of these is forehead, since this region is used to define a sample of surface normals for our pose estimation method. In addition we also study samples of surface normals extracted from the left-cheek and the right-cheek. We selected the forehead and cheeks because these regions on the face are reasonably flat, and hence the associated distributions of surface normals are not unduly dispersed. However, it must be stressed that we do not use the cheeks for pose estimation. We simply use them here to illustrate the consistency of the rotation of the surface normal as the facial pose varies. The regions are taken from the face sequence shown in Fig. 5 above. The surface normal

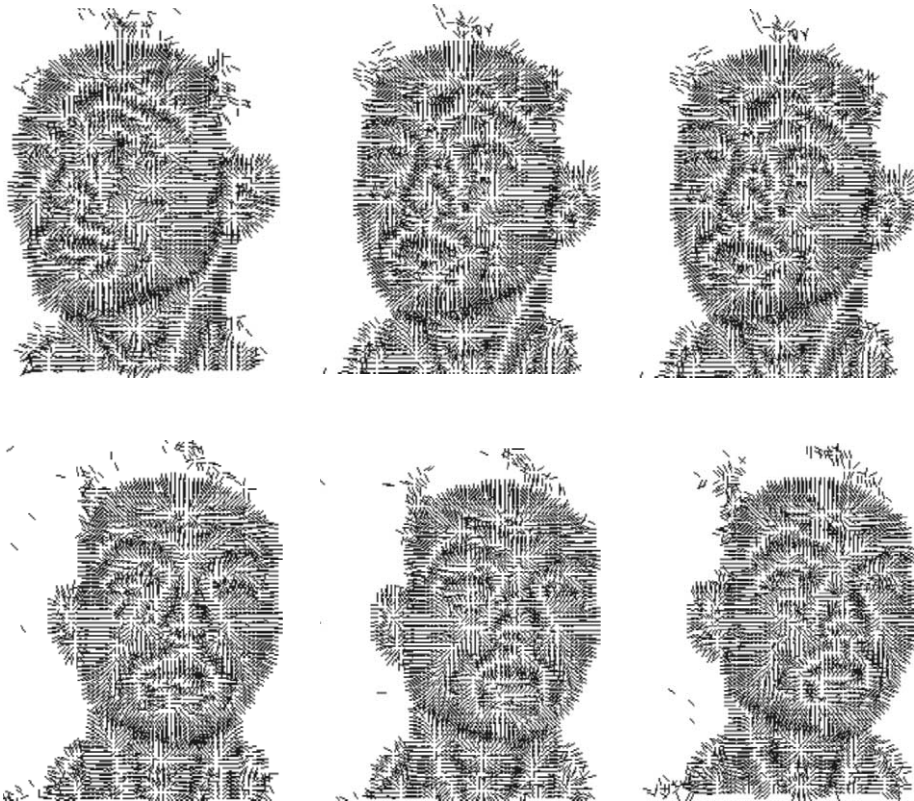


Fig. 6. Needle-maps for a series of views of a real head. The camera direction is approximately fronto-parallel.

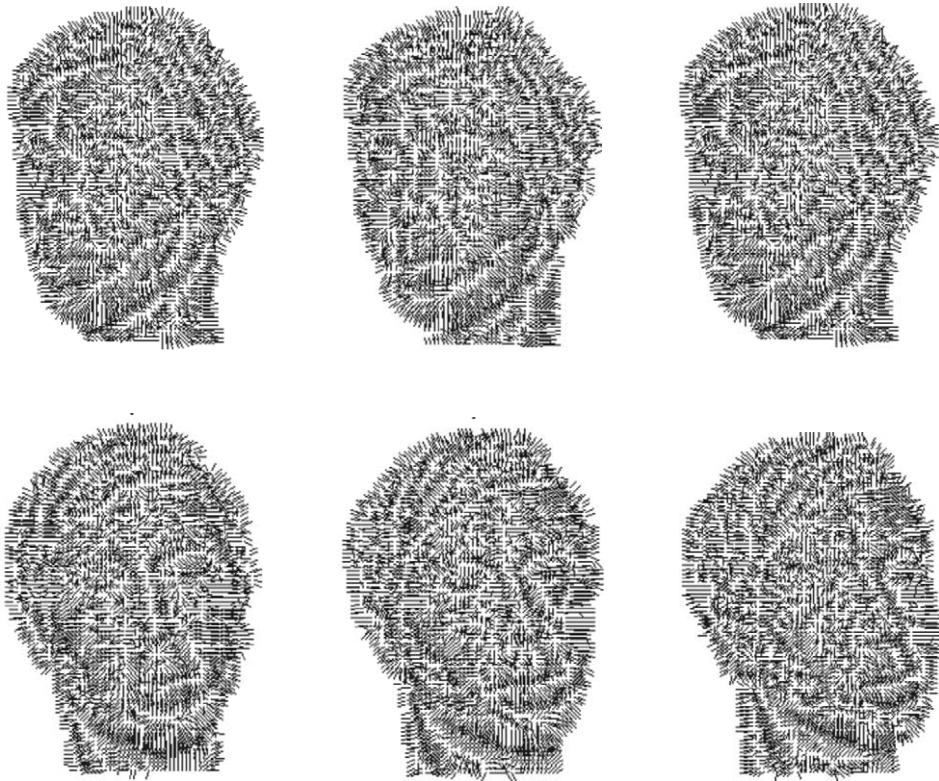


Fig. 7. Needle-maps for a series of views of a model-head. The camera direction of a plaster bust is approximately fronto-parallel.

distributions are displayed in spherical polar form. In each plot we show a unit sphere with the intersections of the surface normal directions marked with dots. The different plots are for samples of surface normals from different facial regions. The top row is for the forehead, the middle row for the left-cheek and the bottom row is for the right-cheek. The distributions in the left-most column are obtained when the head is rotated to the left (i.e., counter-clockwise), the centre column is obtained when the head is fronto-parallel and the right-most column is obtained when the head is rotated right (i.e., clockwise).

There are two features to note from these plots. First, the surface normal distributions for the different regions are expected to form distinct clusters in the polar distribution plots. Second the clusters move in a consistent manner when the head is rotated. In particular, the distributions show that the forehead cluster is compact and rotates in a

manner which is consistent with the movement of the head. The clusters for the cheeks, on the other hand, become more dispersed as the head rotates. This renders them less suitable than the forehead for pose estimation.

#### 4.2. Pose estimation results

Our aim in this section is to investigate whether surface normal orientation histograms can be used for facial pose estimation. Based on the evidence provided in the previous section we focus our attention on the histogram for the forehead region.

In Fig. 9 we show surface normal orientation histograms for the plaster bust as it undergoes rotation between  $-30^\circ$  and  $30^\circ$  in  $10^\circ$  increments. The angle shown in these histograms is the slant angle of the surface normals (i.e.,  $\alpha = \arctan p$ ). In these figures the surface normals are orientated at angles between  $30^\circ$  and  $150^\circ$  to the camera direc-

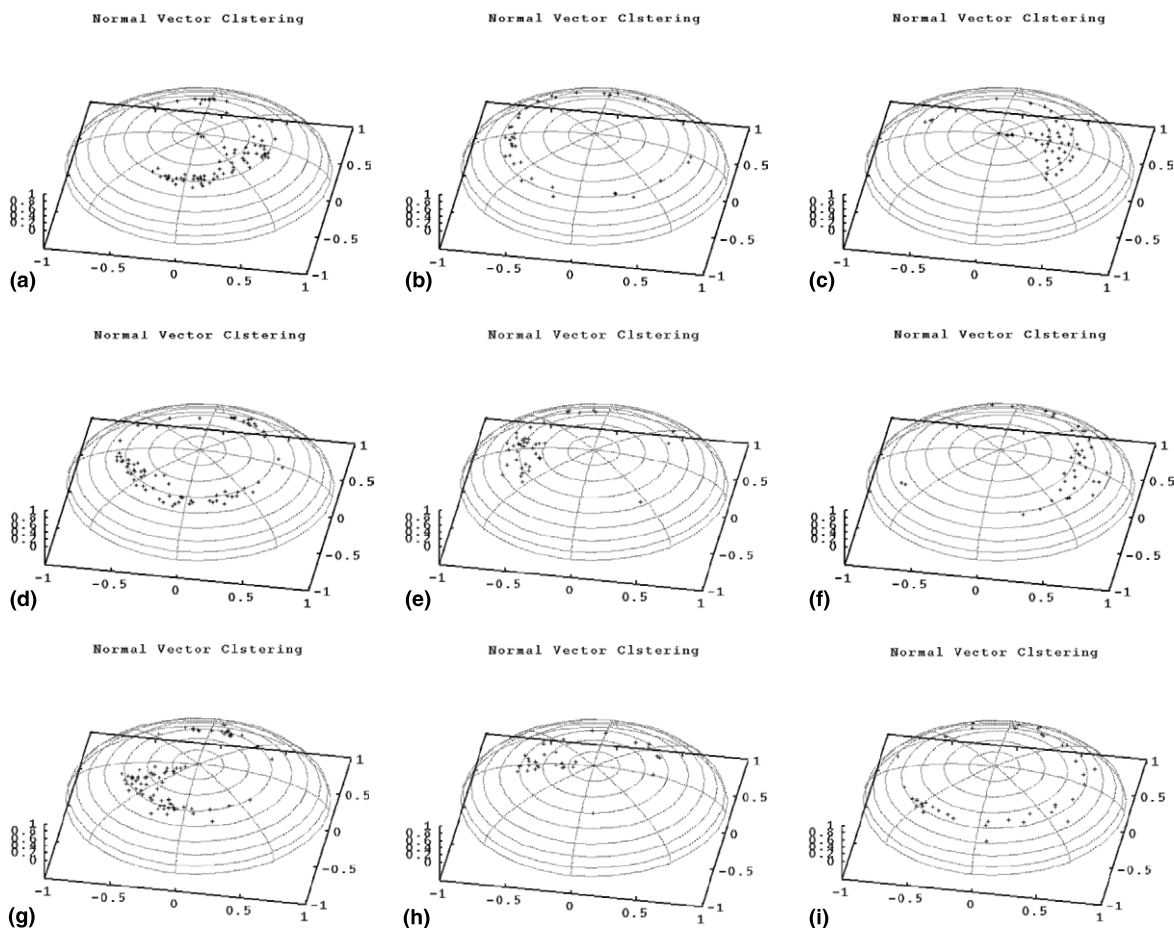


Fig. 8. Normal vector clustering over regions in face. The facial poses are listed from left-to-right. Forehead, left-cheek, and right-cheek are listed from top to bottom: (a) left pose-forehead; (b) front pose-forehead; (c) right pose-forehead; (d) left pose-left cheek; (e) front pose-left cheek; (f) right pose-left cheek; (g) left pose-right cheek; (h) front pose-right cheek; (i) right pose-right cheek.

tion. The main feature to note is that the histograms become increasingly asymmetric as the head rotates further to the left (counter-clockwise) or to the right (clockwise). In Fig. 10 we show the result of transforming the fronto-parallel histogram according to the model described in Section 3. The general shape of the transformed histograms is in reasonable gross agreement with those shown in Fig. 9. In particular the growth of the large peaks at the extremes of the angular distributions is modelled in a way which is qualitatively consistent with the experimental data. The similarity of the two sets of histograms suggests that we can use them as a basis for pose estimation.

We use the plaster bust sequence to evaluate the accuracy of the pose estimates obtained using the similarity of the transformed histograms. In Fig. 11 for the fronto-parallel histogram and in Fig. 12 for the flat histogram, we show the two measures used to compare the histograms. These are the Bhattacharyya distance which measures the similarity of the histograms and should hence take on its maximum value at the optimal pose, and the Euclidean distance which measures the distance between histograms and should hence take on its minimum value for the optimal pose parameters. We plot the two measures as the angle  $\theta$  is varied. The different curves are for the different poses of

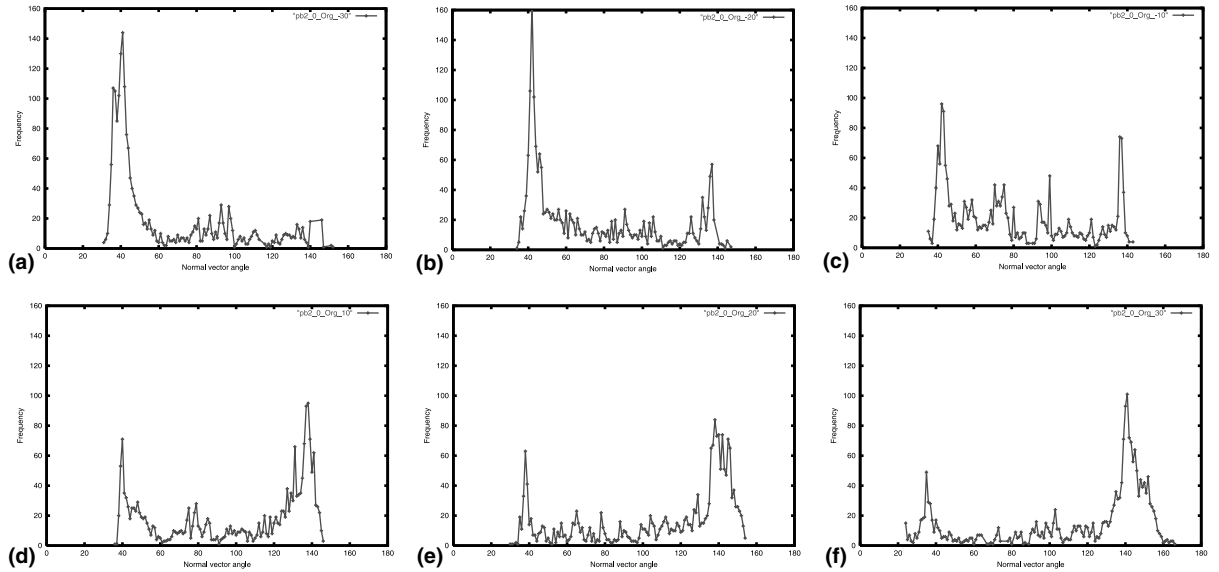


Fig. 9. Histograms of the surface normal slant angles obtained from the forehead of a model-head: (a) pose:  $-30^\circ$ ; (b) pose:  $-20^\circ$ ; (c) pose:  $-10^\circ$ ; (d) pose:  $10^\circ$ ; (e) pose:  $20^\circ$ ; (f) pose:  $30^\circ$ .

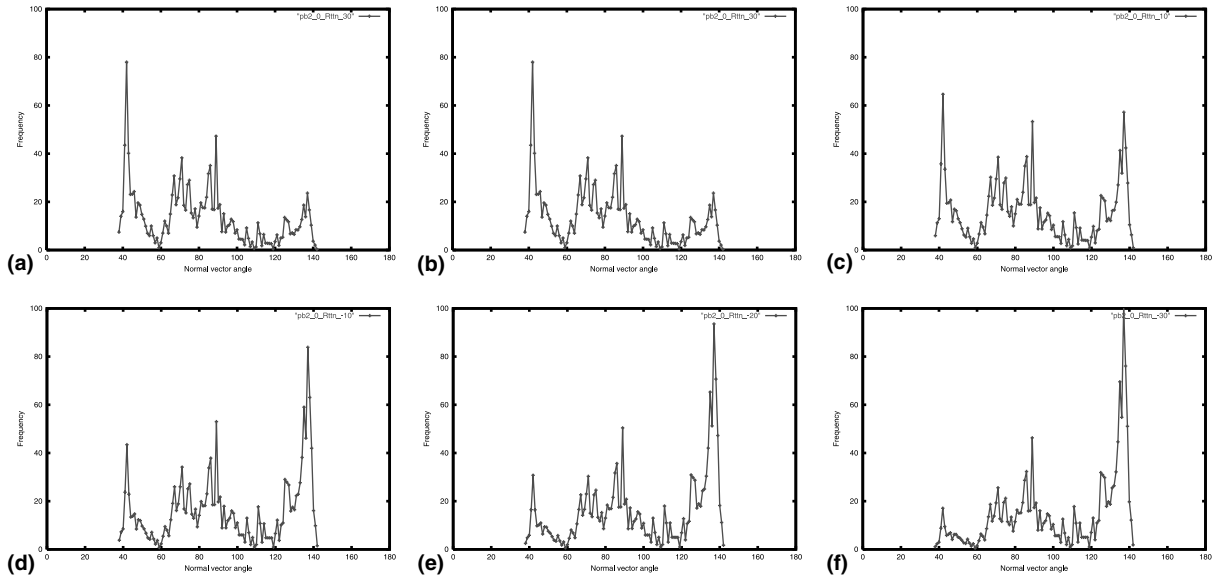


Fig. 10. Transformed histograms: (a) pose:  $-30^\circ$ ; (b) pose:  $-20^\circ$ ; (c) pose:  $-10^\circ$ ; (d) pose:  $10^\circ$ ; (e) pose:  $20^\circ$ ; (f) pose:  $30^\circ$ .

the model. The curves on the plot are for poses from  $-30^\circ$  to  $+30^\circ$  in  $10^\circ$  increments. The two plots in Fig. 11 are obtained when a fronto-parallel image of the plaster bust is used to compute the reference histogram. Plot (a) shows the value of

the Euclidean distance, while plot (b) shows the value of the Bhattacharyya coefficient. The curves in plot (a) are fairly erratic in their variation while those in plot (b) are considerably smoother. The two plots in Fig. 12 show the angular variation in

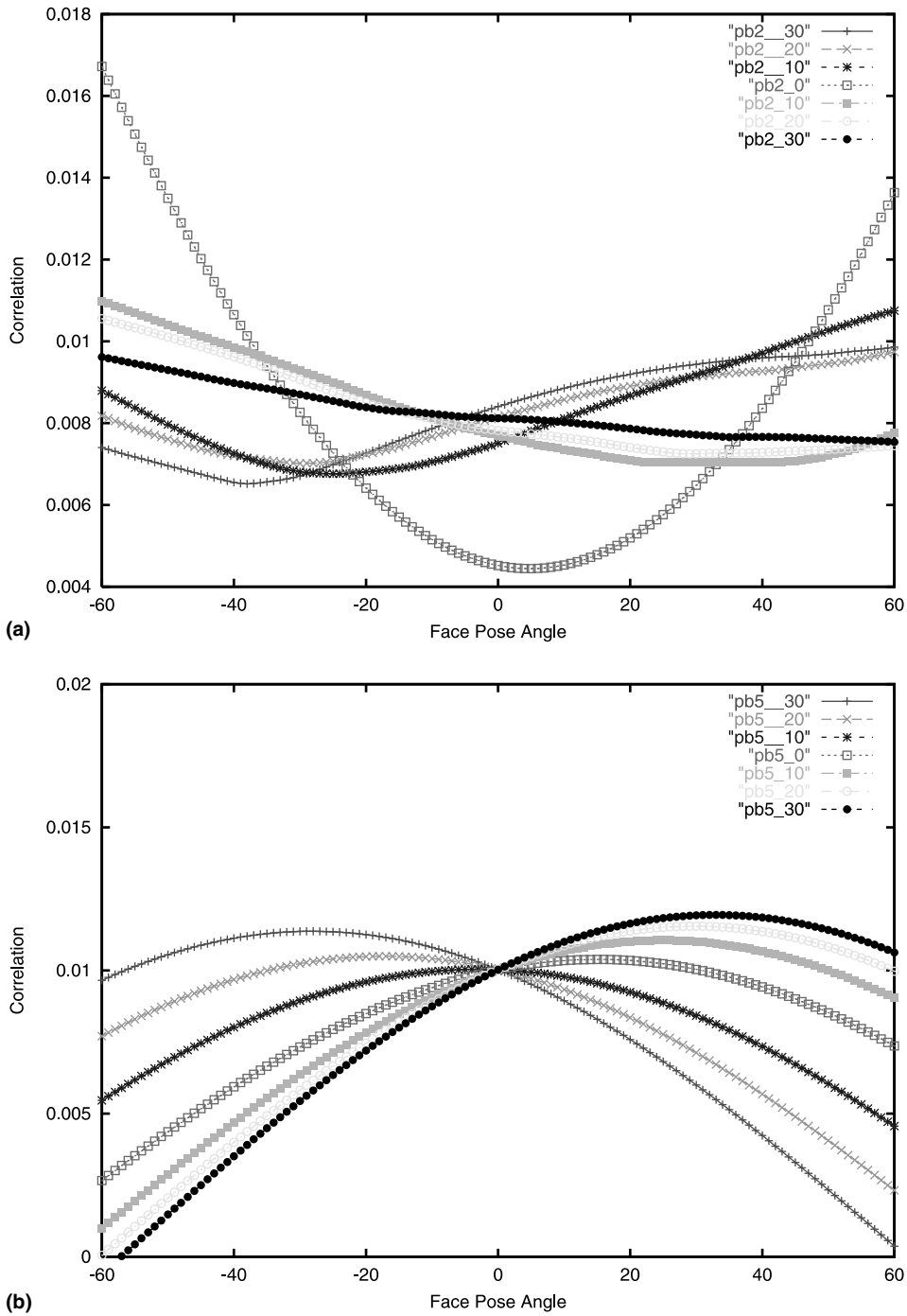


Fig. 11. Measures used to compare the transformed frontal-histogram for the model-head: (a) Euclidean distance versus rotation angle for the frontal histogram; (b) Bhattacharyya co-efficient versus rotation angle for the frontal histogram.

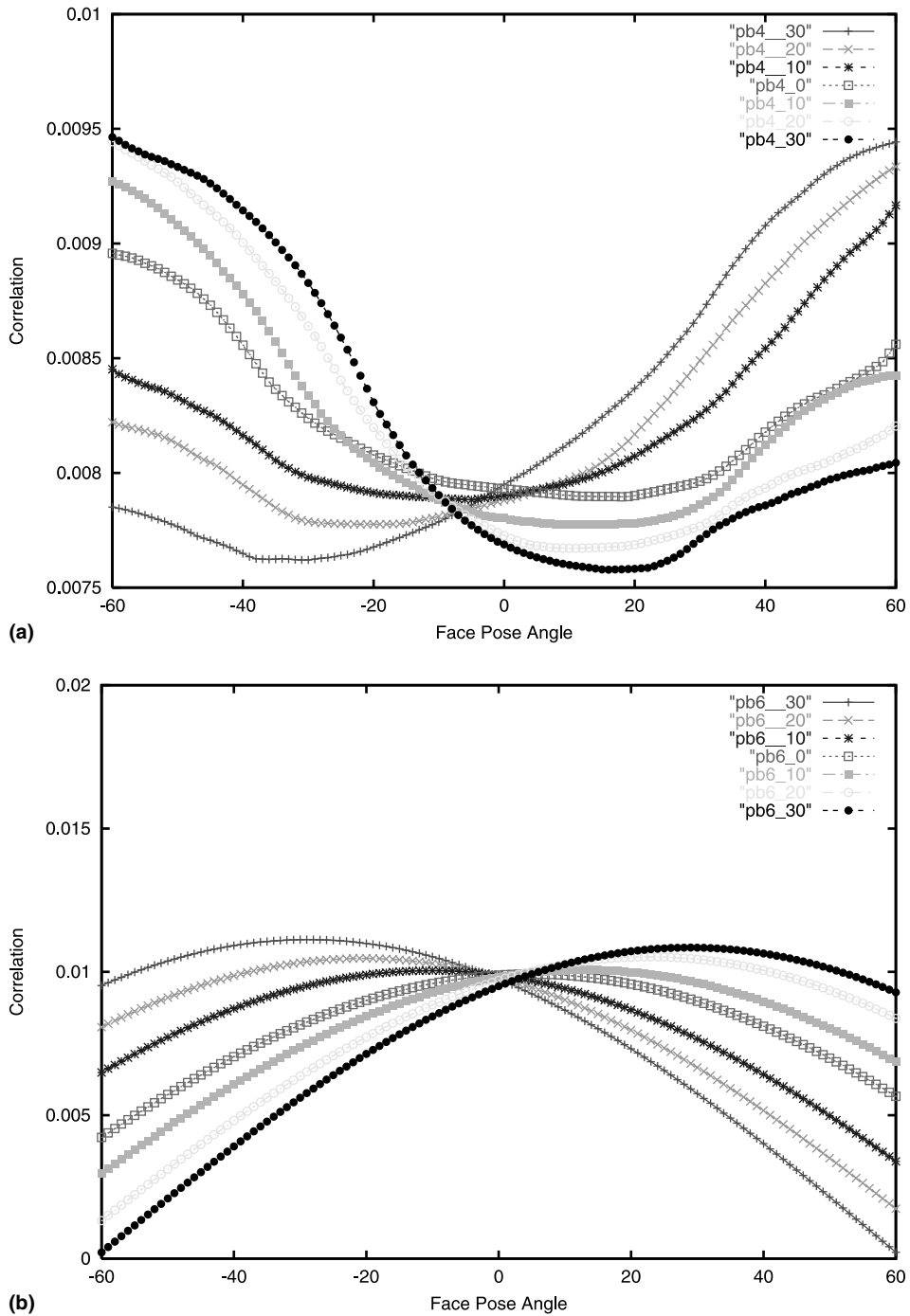


Fig. 12. Measures used to compare the flat histogram for a model-head: (a) Euclidean distance versus rotation angle for the uniform (i.e., flat) histogram; (b) Bhattacharyya co-efficient versus rotation angle for the uniform (i.e., flat).

Table 1  
Ground-truth and estimated facial pose angles for a plaster bust

Group of face	Fronto-parallel histogram		Flat histogram		Ground-truth
	Euclidean	Bhattacharyya	Euclidean	Bhattacharyya	
Face 1	-38	-28	-30	-29	-30
Face 2	-30	-17	-21	-20	-20
Face 3	-25	-3	-5	-10	-10
Face 4	5	15	18	5	0
Face 5	37	25	9	13	10
Face 6	29	30	11	23	20
Face 7	60	33	16	29	30

the Euclidean distance and the Bhattacharyya coefficient when a uniform reference histogram is used. The main features to note from these plots are as follows. In the case of the Bhattacharyya coefficient, the maxima are better defined than the minima of the Euclidean distance. For this reason, we chose to use the Bhattacharyya coefficient as our histogram similarity measure.

Table 1 gives the estimated rotation angles for each of the images of the plaster bust shown in Fig. 5. The plots of the histogram similarity measure are shown in Figs. 11 and 12. The results obtained by using the fronto-parallel histogram and the Euclidean distance lead to pose angles that are over-estimated. The pose angles of the fronto-parallel histogram using the Bhattacharyya coefficient are not accurately determined. For the flat histogram, the Euclidean distance gives under-estimated angles for the left-most two poses (*{Face 6}* and *{Face 7}*). The pose estimates obtained using the Bhattacharyya coefficient have an error of less than  $3^\circ$  when they are compared to ground-truth. The main conclusion to be drawn from this table is that the most consistently accurate parameters are recovered by maximising the Bhattacharyya coefficient with the uniform histogram. The alternative methods all perform less well.

Table 2 reports results for natural images. For the eight subjects, we report the mean and standard deviation of the pose angle for each of the seven poses and the three different facial attitudes (tilted up, facing forward and tilted down). The pose angles are obtained using the Bhattacharyya coefficient and the uniform histogram (which produced the best result in Table 1). The angles decrease in a regular way as the head pose rotates.

Table 2  
Mean and standard deviation of the estimated facial pose angle for each group of 21 facial images

Pose group	Mean	Standard deviation
Pose 1	27.17	1.80
Pose 2	21.04	1.55
Pose 3	12.75	1.19
Pose 4	0.33	0.95
Pose 5	-9.08	1.18
Pose 6	-14.92	1.30
Pose 7	-22.46	1.33

The average interval between the mean pose angles is  $8.3^\circ$ . The standard deviations of the pose angles for the different poses are relatively small (less than  $1.5^\circ$ ). This would suggest that the method can be used to make rough pose estimates with an accuracy of a few degrees.

## 5. Conclusions

We have reported the application of a new shape-from-shading technique to the problem of facial pose estimation. The pose estimation process relies on histogram correlation to recover the facial rotation angles. The method can yield pose estimates that are accurate up to a few degrees. We experimented with pose estimation on a plaster bust and a set of facial images of various subjects. There are two choices which appear to improve algorithm performance. First, the use of the Bhattacharyya coefficient as a histogram similarity measure is superior to the use of the Euclidean distance. Second, the best results were obtained with a uniformly populated reference histogram.

These observations apply to experiments with both a plaster bust and with natural face images.

There are a number of limitations of the proposed method. First, if the reference histogram is to be computed using a fronto-parallel view of the subject then (a) a different histogram may be necessary for each subject and (b) the subjects head must be accurately aligned in the correct pose. Second, at present the parallelogram region on the forehead is selected by hand. Finally, the method is only effective when the head is rotated by less than 40° from the frontal pose. However, the advantage of the method is that it does not rely on detailed feature alignment to a 3D model template.

## References

- Atick, J., Griffin, P., Redlich, A., 1996. Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. *Neural Computation* 8 (6), 1321–1340.
- Belhumeur, P., Hespanha, J., Kriegman, D., 1997. Eigenfaces vs. fisherfaces: Recognition using class-specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.* 19, 711–720.
- Belhumeur, P., Kriegman, D., 1996. What is the set of images of an object under all possible lighting conditions. *Computer Vision and Pattern Recognition* 96, 270–277.
- Braje, W., Kersten, D., Tarr, M., Troje, N., 1998. Illumination effects in face recognition. *Psychobiology* 26 (4), 371–380.
- Cootes, T., Taylor, C., Lanitis, A., 1994. Active shape models: Evaluation of a multi-resolution method for improving image search. In: *Proc. British Machine Vision Conference*, pp. 327–336.
- Devijver, P., Kittler, J., 1982. *Pattern Recognition – A Statistical Approach*. Prentice-Hall, Englewood Cliffs.
- Enns, J., Shore, D., 1997. Separate influences of orientation and lighting in the inverted-face effect. *Perception Psychophys.* 59 (1), 23–31.
- Gee, A., Cipolla, R., 1994. Estimating gaze from a single view of a face. *Internat. Conf. on Pattern Recognition* 94 A, 758–760.
- Georgiades, A., Kriegman, D., Belhumeur, P., 1998. Illumination cones for recognition under variable lighting: Faces. *Computer Vision and Pattern Recognition* 98, 52–59.
- Hattori, K., Matsumori, S., Sato, Y., 1998. Estimating pose of human face based on symmetry plane using range and intensity image. *Internat. Conf. on Pattern Recognition* 98, 1183–1187.
- Huang, J., Li, D., Shao, X., Wechsler, H., 1997. Pose discrimination and eye detection using support vector machines. In: *Face Recognition, From Theory to Applications – NATO ASI Series, Series F: Computer and Systems Sciences*, Vol. 163. Springer, Berlin, pp. 528–535.
- Huang, J., Shao, X., Wechsler, H., 1998. Face pose discrimination using support vector machines. *Internat. Conf. on Pattern Recognition* 98, 154–156.
- Jacobs, D., Belhumeur, P., Basri, R., 1998. Comparing images under variable illumination. *Computer Vision and Pattern Recognition* 98, 610–617.
- Kemp, R., Pike, G., White, P., Musselman, A., 1996. Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues. *Perception* 25 (1), 37–52.
- McKenna, S., Gong, S., Collins, J., 1996. Face tracking and pose representation. In: *British Machine Vision Conference* 96, pp. 755–764.
- Romdhani, S., Gong, S., Psarrou, A., 1999. Multi-view nonlinear active shape model using kernel PCA. *British Machine Vision Conference* 99, 483–492.
- Troje, N., Bulthoff, H., 1996. Face recognition under varying poses: The role of texture and shape. *Vision Res.* 36 (12), 1761–1771.
- Wiskott L., Fellous, J., Kruger, N., Malsburg, C., 1996. Face recognition by elastic bunch graph matching. *Tech. Rep. IRINI 96-08*, Ruhr-Universität Bochum.
- Worthington, P., Hancock, E., 1999. New constraints on data-closeness and needle-map consistency for shape-from-shading. *IEEE Trans. Pattern Anal. Machine Intell.* 21, 1250–1267.
- Yuille, A., Yuille, M.F., Zhang, T., 1998. Image warping for shape recovery and recognition. *Computer Vision and Image Understanding* 72 (3), 351–359.