

Facial Pose using Shape-from-Shading

Kwang Nam Choi, Philip Worthington and Edwin R. Hancock
Department of Computer Science
University of York, York, YO10 5DD, UK

Abstract

This paper reports the application of a novel shape-from-shading technique to estimating 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 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 to within a few degrees.

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 [1, 2, 4, 8, 11, 14]. The aim in this paper is to investigate whether shape-from-shading can be used for the purposes estimating facial pose. Because of its perceived fragility, the use of needle map information has received little interest. 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 recover locally consistent needle-maps. 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 focussed on developing more sophisticated physically based models to account for the observed illumination. For example Jacobs, Belhumeur and Basri [6] have used multiple light-sources to extract shape-information.

In a recent paper we have reported an improved framework for shape-from-shading which overcomes some of the shortcomings listed above [13]. Our 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, we have developed alternatives to the bland 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 the current work is to investigate whether the new shape-from-shading technique can be used for recovering facial pose. The idea 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 orientation angles changes under horizontal and vertical rotations. The pose is recovered by locating the rotation angles that lead to maximum similarity between the observed orientation histogram and the transformed model histogram. 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 distance and a uniform initial distribution of orientations is assumed.

2 Shape-from-Shading

Central to shape-from-shading is the idea that local regions in an image $E(x, y)$ correspond to illuminated patches of a piecewise continuous surface whose height is represented by the function $z = f(x, y)$. The observed image intensity $E(x, y)$ will vary depending on the material properties of the surface (i.e whether it is matte or specular), the orientation of the surface at the co-ordinates (x, y) , and the direction of illumination. The *reflectance map*, $R(p, q)$, provides an explicit mapping between patch orientation and brightness, in terms of the surface gradient in the x and y direction, i.e. $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$.

Shape-from-shading aims to recover the needle-map or local surface normal distribution $\mathbf{n}(x, y)$ from the image-data $E(x, y)$. Here we consider Lambertian surfaces which have a matte appearance and reflect incident light uniformly in all directions. As a result the light reflected is simply proportional to the orientation of the patch relative to the light source direction. Suppose that the local surface normal is $\mathbf{n} = (-p, -q, 1)^T$. For a light source at infinity, we can similarly write the source direction as $\mathbf{s} = (-p_l, -q_l, 1)^T$. For a Lambertian surface, the reflectance map is given by $R(p, q) = \mathbf{n} \cdot \mathbf{s}$. The image irradiance equation (IIR)[7] states that the measured image brightness is proportional to the radiance at the corresponding surface point. Normalising both the image intensity and the reflectance map, the IIR is simply $E(x, y) = R(p, q)$.

We have recently developed a new framework for shape-from-shading. The new method guarantees data-closeness by treating the IIR as a hard constraint. In other words, we aim to recover a valid needle-map which satisfies the IIR at every iteration. Subject to this data-closeness constraint, the task of shape-from-shading becomes that of iteratively improving the organisation of the needle-map using curvature consistency constraints.

Our approach is a geometric one. We view the IIR as defining a cone of ambiguity about the light source direction for each surface normal. The individual surface normals which constitute the needle-map can only assume directions the fall on this cone. At each iteration the updated normal is free to move away from the cone under the action of the local consistency constraints. However, it is subsequently mapped back onto the closest normal residing on the cone. By applying this constraint, we gain dual advantages in terms of both numerical stability and obviating the need for a Lagrange multiplier. More importantly, the needle-map evolves via a series of intermediate states which are each solutions of the IIR.

2.1 Hard Constraints

The new framework requires us to minimize the constraint functional

$$I = \int \int \psi(\mathbf{n}(x, y), \mathbf{N}(x, y)) dx dy \quad (1)$$

whilst satisfying the hard constraint, imposed by the image irradiance equation

$$\int \int (E(x, y) - \mathbf{n}(x, y) \cdot \mathbf{s}) dx dy = 0 \quad (2)$$

Here $\mathbf{N}(x, y)$ is the set of local neighbourhood normals about the image location (x, y) . The quantity $\psi(\mathbf{n}(x, y), \mathbf{N}(x, y))$ is a localised constraint function based on the current surface normal estimates. The size of the neighbourhood system $\mathbf{N}(x, y)$ may be adjusted in the light of the constraint function ψ .

We impose the hard constraint after each iteration by mapping the intermediate update, in this case $\bar{\mathbf{n}}_{i,j}^k$, back to the most similar normal satisfying the IIR. The update is realised by rotating the current average surface normal using the rotation matrix Θ . The update equation is

$$\mathbf{n}_{i,j}^{k+1} = \Theta \bar{\mathbf{n}}_{i,j}^k \quad (3)$$

The matrix Θ rotates the average normal vector by an amount equal to the difference between the apex angle of the cone, and the angle between the intermediate normal and the light source direction, i.e. by $\theta = \cos^{-1} E - \phi$ where $\phi = \frac{\bar{\mathbf{n}}_{i,j}^k \cdot \mathbf{s}}{\|\bar{\mathbf{n}}_{i,j}^k\| \cdot \|\mathbf{s}\|}$. The axis of rotation is $(u, v, w)^T = \bar{\mathbf{n}}_{i,j}^k \times \mathbf{s}$, which is perpendicular to both the light source direction and the intermediate update. Hence, the rotation matrix is

$$\Theta = \begin{pmatrix} c + u^2 c' & -ws + uv c' & vs + uw c' \\ ws + uv c' & c + v^2 c' & -us + vw c' \\ -vs + uw c' & us + vw c' & c + w^2 c' \end{pmatrix}$$

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

2.2 Initialisation

The new framework requires an initialisation which ensures that the IIR is satisfied from the outset. In other words we must choose a normal direction from the infinite possibilities which define the cone of ambiguity. We select the initial surface normals such that their projections onto the image plane are aligned in the direction opposite to that of the local image gradient. This results in an initialisation with an implicit bias towards convex rather than concave surfaces. In other words, bright regions are assumed to correspond to peaks.

2.3 Curvature Consistency Constraints

The use of curvature consistency in SFS was proposed by Ferrie and Lagarde [5] as a non-iterative post-processing step. Specifically, they use the global consistency of the principal curvatures [10] to refine the surface estimate returned by local shading analysis. Here, we attempt to formulate curvature consistency as a *replacement* for the quadratic

smoothness constraint. To avoid multiple objective measures in the constraint function, we seek a single measure of topographic class. An obvious candidate here is to use the $H - K$ labels which are based on the signs and zeros of the mean and Gaussian curvatures. The assignment of these labels can be quite cumbersome since it requires the setting of four curvature thresholds. For this reason we choose to base our measure of curvature consistency on the shape index of Koenderink and van Doorn [9]. This is a single continuous measure which encodes the same topographic class information as $H - K$ labels in an angular representation, without the need to set thresholds.

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

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

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

The eigenvalues of the Hessian matrix, found by solving the equation $|\mathcal{H} - \kappa \mathbf{I}| = 0$, are the principal curvatures of the surface, denoted $\kappa_{1,2}$. The shape index is defined in terms of the principal curvatures

$$\phi = \frac{2}{\pi} \arctan \frac{\kappa_2 + \kappa_1}{\kappa_2 - \kappa_1} \quad \kappa_1 \geq \kappa_2 \quad (5)$$

and may be expressed in terms of surface normals thus

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

Figure 1 shows the range of shape index values, the type of curvature which they represent, and the grey-levels used when displaying shape-index values.

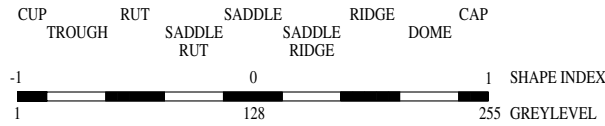


Figure 1: The shape index scale ranges from -1 to 1 as shown. The shape index values may be encoded as grey-level values.

We use curvature consistency to control the robust weighting kernel applied to the variation in the needle-map direction. The idea is a simple one. We use the variance of the shape-index in the neighbourhood N to control the width σ of the robust error-kernel applied to the directional derivatives of the needle map. The kernel width determines the level of smoothing applied to the surface normals in the neighbourhood. If the variance of the shape index is large i.e. the neighbourhood contains a lot of topographic structure, then we choose a small kernel width. This limits the local smoothing and allows significant local variation in the local needle-map direction. From a topographic viewpoint, we

can see the rationale for this choice by considering the behaviour of the needle-map and the shape-index at ridges and ravines. For such features, the direction of the needle-map changes rapidly in a particular direction. These two structures are parabolic lines which intercede between elliptic and hyperbolic regions. As a result there is a rapid local variation in shape index. Turning our attention to the case where the shape-index variance is small, then the kernel width is large. This is the case when we are situated in a hyperbolic or elliptic region. Here the shape-index is locally uniform and the needle-map direction varies slowly.

To capture these requirements we draw on the robust approach to SFS recently described in [12]. Specifically, we define a smoothness constraint on the directional derivatives of the surface normals and let

$$\psi(\mathbf{n}, \mathbf{N}) = \rho_\sigma \left(\left\| \frac{\partial \mathbf{n}}{\partial x} \right\| \right) + \rho_\sigma \left(\left\| \frac{\partial \mathbf{n}}{\partial y} \right\| \right) \quad (7)$$

where ρ_σ is a robust kernel with adaptive width

$$\sigma = \sigma_0 \exp \left(- \left(\frac{1}{N} \sum_{l \in \mathbf{N}} \frac{(\phi_l - \phi_c)^2}{\delta \phi_d^2} \right)^{\frac{1}{2}} \right) \quad (8)$$

Here, ϕ_c is the shape index associated with the central normal of the neighbourhood and $\delta \phi_d$ is the difference in shape index between the centre values of adjacent curvature classes. Using the scale of Figure 1, $\delta \phi_d = \frac{1}{8}$. Applying the calculus of variations, discretizing and re-arranging, yields the general update equation

$$\mathbf{n}_{i,j}^{(k+1)} = \Theta \left(\frac{\partial}{\partial x} \rho'_\sigma \left(\left\| \frac{\partial \mathbf{n}}{\partial x} \right\| \right) + \frac{\partial}{\partial y} \rho'_\sigma \left(\left\| \frac{\partial \mathbf{n}}{\partial y} \right\| \right) \right) \quad (9)$$

In [12] we demonstrated that the best performance was obtained with a continuous version of Huber's M-estimator. The error-kernel is defined on the residual η and is given by

$$\rho_\sigma(\eta) = \frac{\sigma}{\pi} \log \cosh \left(\frac{\pi \eta}{\sigma} \right) \quad (10)$$

If the difference in shape index values over the neighbourhood is large, a small value of σ results. This narrows the kernel so that it saturates to produce a heavy smoothing effect. In contrast, when the differences are small, the kernel is widened so that the estimate of the centre normal is not modified.

3 Pose Estimation

In this section we describe how the needle-maps delivered by the 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

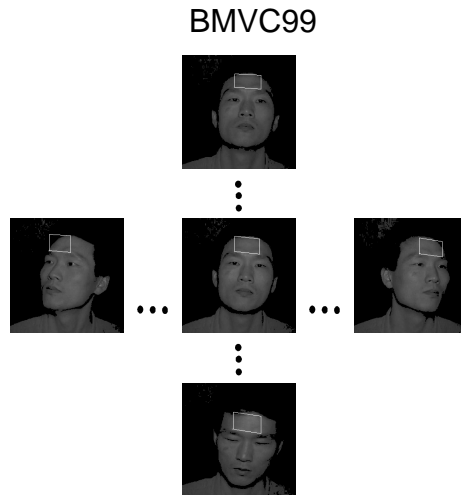


Figure 2: Window used for pose estimation.

recovered surface normal 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 horizontal rotation of the surface normal about the image plane normal. This angle is equal to $\alpha = \arctan p$. The second angle is the vertical rotation about the image plane surface normal $\beta = \arctan q$. These two angles are binned in a two-dimensional histogram for a sample of surface normals. The normalised contents for the bin-centred on the rotation angles α and β is denoted by $h(\alpha, \beta)$.

3.2 Surface Normal Sample

The sample of surface normals used in our pose estimation algorithm is selected from the forehead. The reason for choosing this facial region are twofold. Firstly, it is devoid of topographic structure such as a ridges and ravines. Secondly, it is a approximately planar.

The forehead region is a parallelogram which is selected as follows. The leftmost and rightmost extremes are vertical lines defined by the center-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 the region are extracted using a recently described Fourier-domain matched filter technique [3]. An example of the automatic segmentation of the region-of-interest for different facial poses is shown in Figure 2.

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 θ . The second angle is the vertical rotation ϕ . As the head rotates, then so the histogram bin-contents transforms. There are two factors controlling the sampling of the needle-map 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

$$h_{\theta, \phi}(\alpha + \theta, \beta + \phi) = h_{0,0}(\alpha, \beta) \left| \frac{\sin(\alpha + \theta) \sin(\beta + \phi)}{\sin(\alpha) \sin(\beta)} \right| \quad (11)$$

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 angles θ and ϕ which maximise the similarity between the histogram g 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. Another approach adopted in our experiments is to assume a flat histogram. This is clearly very poor model of the distribution of orientation angles across an approximately planar surface region. However, if the reference histogram is taken as a prior model, then uniformity corresponds to the flat or maximum entropy prior.

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.

$$\begin{pmatrix} \theta \\ \phi \end{pmatrix} = \arg \min_{\theta, \phi} \sum_{\alpha} \sum_{\beta} \left[h_{\theta, \phi}(\alpha, \beta) - g(\alpha, \beta) \right]^2 \quad (12)$$

The second similarity measure is the Bhattacharyya coefficient. According to this similarity criteria the vector of pose parameters is

$$\begin{pmatrix} \theta \\ \phi \end{pmatrix} = \arg \max_{\theta, \phi} \sum_{\alpha} \sum_{\beta} \sqrt{h_{\theta, \phi}(\alpha, \beta) g(\alpha, \beta)} \quad (13)$$

4 Experiments

We commence our experimental study by showing some examples of the needle-maps extracted using our shape-from-shading scheme. We have used both real subjects and a model-head in our experiments. The model head is used to compare results against ground-truth. Figure 3 shows a sequence of images of the head of a real subject in different poses. The corresponding sequence of needle maps are shown in figure 4. There are several features of these needle-maps which deserve further comment. Firstly, the topographic features such as the nose, lips and eye-sockets are well defined. Secondly, these features are stable under varying pose.

To compare the orientation estimates with ground truth we have used a sequence of images of the model-head in which the horizontal rotation is known. Table 1 shows the estimated rotation angles for each of the images of a plaster bust. The main conclusion

BMVC99

	Front Histogram		Flat Histogram		Ground-Truth
	<i>(Euclid.)</i>	<i>(Bhat.)</i>	<i>(Euclid.)</i>	<i>(Bhat.)</i>	
(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

Table 1: Actual \times Estimated Facial Pose Angle for a plaster bust.

	Front Histogram		Flat Histogram		Template
	<i>(Euclid)</i>	<i>(Bhat.)</i>	<i>(Euclid.)</i>	<i>(Bhat.)</i>	
(Face 1)	-35	-16	-11	-16	-23.0
(Face 2)	-31	-22	-18	-19	-16.4
(Face 3)	-12	-14	-8	-11	-7.1
(Face 4)	-2	-3	-7	-2	-4.8
(Face 5)	-1	3	8	7	11.7
(Face 6)	9	13	13	18	28.3
(Face 7)	29	18	13	24	36.4

Table 2: Actual \times Estimated Facial Pose Angle for a real face.

to be drawn from this plot is that the most consistently accurate parameters are recovered by maximising the Bhattacharyya coefficient with the uniform histogram. The alternative methods all perform rather poorly.

Table 2 repeats this analysis for the natural image sequence. The poses of the head are separated by approximately 10 degrees. However, we do not have accurate ground truth. For this reason we compare our pose estimates with those obtained from the template matching algorithm. However, for rotations in excess of 25 degrees, this method returns unreliable angle estimates. The estimated angles again support our claim that the use of Bhattacharyya distance and a uniform histogram offers the best performance.



Figure 3: A series of views of a real-face. The camera direction of views is approximately fronto-parallel.

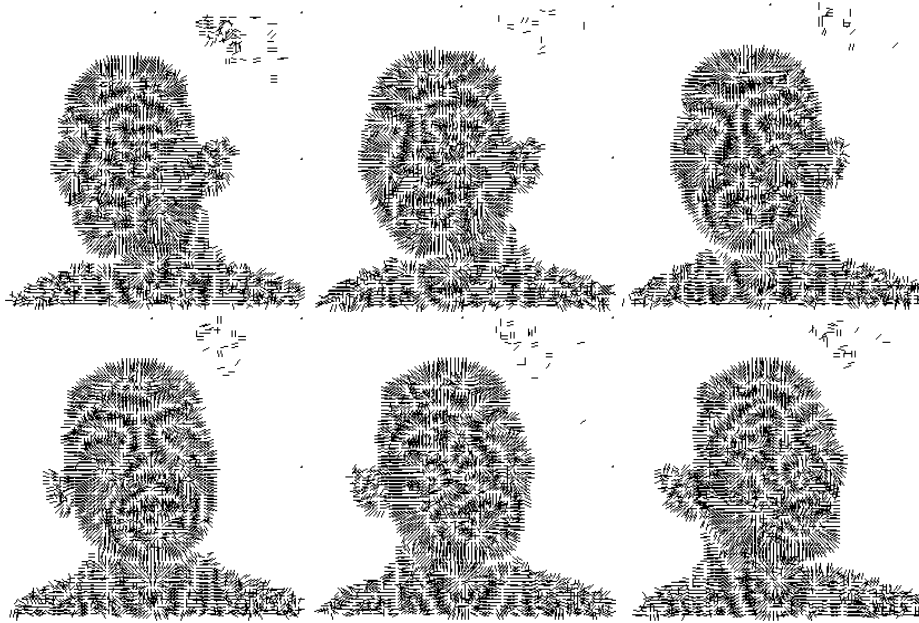


Figure 4: Needle map for a series of views of a real head. The camera direction of views is approximately fronto-parallel.

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 to a few degrees.

This work can be viewed as the first-step in a more ambitious programme of work aimed at using the improved needle-maps for face analysis. We are currently investigating how shape-index spectra extracted from the needle-maps can be used for the purposes of recognising individuals.

References

- [1] Atick J.J., Griffin P.A. and Redlich A.N., "Statistical Approach to Shape from Shading: Reconstruction of Three-Dimensional Face Surfaces from Single Two-Dimensional Images", *Neural Computation* **8**, pp. 1321-1340, 1996.
- [2] Braje W.L., Kersten D., Tarr M.J. and Troje N.F., "Illumination Effects in Face Recognition", *Psychobiology* **26**, pp. 371-380, 1998.
- [3] Choi K.N., Cross A.D.J. and Hancock E.R., "Localising Facial Features with Matched Filters", *First International Conference on Audio- and Video-based Biometric Person Authentication, LNCS, 1206*, pp. 11-20, 1997.

BMVC99

- [4] Enns J.T. and Shore D.I. , “Separate Influences of Orientation and Lighting in the Inverted-Face Effect”, *Perception and Psychophysics* **59**, pp. 23-31, 1997.
- [5] Ferrie, F.P. and Lagarde, J., “Curvature Consistency Improves Local Shading Analysis”, *Proc. IEEE International Conference on Pattern Recognition I*, pp. 70-76, 1990.
- [6] Jacobs, D., Belhumeur, P.N. and Basri, R., “Comparing Images Under Variable Illumination”, *CVPR98*, pp. 610-617, 1998.
- [7] Horn, B.K.P. and Brooks, M.J. (eds.), *Shape from Shading*, MIT Press, Cambridge, MA, 1989.
- [8] Kemp R., Pike G., White P. and Musselman A., “Perception and Recognition of Normal and Negative Faces: The Role of Shape from Shading and Pigmentation Cues”, *Perception* **25**, pp. 37-52, 1996.
- [9] Koenderink J. and van Doorn A., “Surface Shape and Curvature Scales,” *Image and Vision Computing* **10**, pp. 557-565, 1992.
- [10] Parent P. and Zucker S., “Curvature Consistency and Curve Detection,” *J. Opt. Soc. America A*, **2**, pp. 5, 1985.
- [11] Troje N.F. and Bulthoff H.H., “Face Recognition under Varying Poses: The role of Texture and Shape”, *Vision Research* **36**, pp. 1761-1771, 1996.
- [12] Worthington P. and Hancock E., “Needle Map Recovery using Robust Regularizers,” *Proc. British Machine Vision Conf. BMVA Press*, **I**, pp. 31-40, 1997.
- [13] Worthington P.L. and Hancock, E.R., “Data driven shape-from-shading using Curvature Consistency Constraints”, *CVPR99*, to appear.
- [14] Yuille A.L., Ferraro M. and Zhang T., “Image Warping for Shape Recovery and Recognition”, *Computer Vision and Image Understanding* **72**, pp. 351-359, 1998.