Facial Shape Estimation in the Presence of Cast Shadows

William A. P. Smith and E. R. Hancock
Department of Computer Science,
The University of York, UK
{wsmith,erh}@cs.york.ac.uk

Abstract

This paper describes a method for cast shadow removal from obliquely illuminated images of faces. The method draws on a statistical model of surface normal directions. The model is fitted to shadowed facial images using robust statistics and constraints provided by shape-from-shading. Regions associated with poor fit residuals are associated with shadow regions. We illustrate the method on the Yale B database where it gives both good shadow map estimates and fills-in the facial surface in the shadow regions.

1 Introduction

The presence of cast shadows frustrates a number of face analysis tasks. Simple local illumination models employed by many of the most promising face analysis approaches (e.g. photometric stereo [5] and spherical harmonic images [2]) fail to account for the effect of cast shadows. It is frequently assumed that gallery images are taken under close-to-frontal lighting and the effects of cast shadows can be ignored. However, if only a single gallery image exists and it contains cast shadows, their effect must be dealt with. The problem of cast shadow estimation and surface completion for these cast shadow regions is therefore clearly an important one.

When a surface is illuminated by a single point light source, a point on the surface is *in shadow* when it is not visible from the light source. In other words, no light reaches the point and the measured intensity at the pixel which corresponds to the point is zero. In this paper we restrict ourselves to single point light sources, since extended or multiple light sources produce more complicated effects.

There are two scenarios which result in a point being in shadow and these are modeled in quite different ways. An *attached shadow* (also called a self-shadow) occurs when a point on the surface is oriented away from the light source, thereby occluding itself from the illumination. Shadows of this sort are easily modeled, since they depend only upon the local geometry of the surface (the normal direction). On the other hand, *cast shadows* are caused when an entirely different region of the surface intersects the path from the light source to the point in question. Cast shadows are more complex to model since they are dependent on the global geometry of the surface. In Figure 1 we show an artificial example to illustrate the difference between the two types of shadow. We render a surface of Gaussian peaks with a single light source from behind the peaks to the top right. The non-shadow regions have been shaded red with Lambertian reflectance. The regions of

BMVC 2006 doi:10.5244/C.20.80

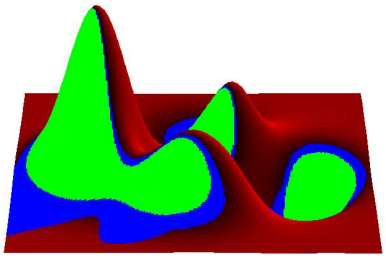


Figure 1: Example of cast and attached shadows.

attached shadow are shown in green, while the regions shown in blue are cast shadows caused by the peaks intercepting the light source.

For a given illumination condition, the regions of a surface (whose height function is known) that lie in cast shadow may be calculated using ray-tracing. With only one intensity image as input, the situation becomes more difficult. Inter-reflections and the presence of ambient illumination mean that shadow pixels in real world images will not necessarily be measured as having zero intensity. One approach is to estimate the 3D shape from the image intensity and calculate the cast shadow regions from the estimated shape. In non-shadow regions, the measured intensity conveys information about the local surface orientation. Recovering this orientation from a single intensity image is the shape-from-shading problem [10]. However, to recover the surface height function (which is necessary to identify cast shadow regions) the surface orientation at every point is required. Unfortunately, shadow regions convey no information about facial shape or texture. The result is a 'chicken or egg' situation: accurate knowledge of global shape is required to estimate cast shadow regions, but cast shadow regions disrupt the accurate estimation of the global shape from image intensity.

In this paper we suggest an iterative approach to this problem which uses hard irradiance constraints in non-shadow regions to recover accurate shape, whilst using a class-based statistical model to estimate the shape of shadowed regions. The method is able to recover a surface which accurately reproduces the cast shadows present in the input image, whilst using class-based surface completion to recover information about the face in these shadowed regions.

2 Facial Shape-from-shading

In recent work Smith and Hancock [7] presented a novel approach to facial shape-from-shading which combined a hard irradiance constraint with a statistical model that captures variations in surface normal direction. Under conditions of close-to-frontal lighting this approach provides a good estimate of the facial shape from a single image. However, the method breaks down as lighting becomes more extreme, conditions in which cast shadows become significant and a local shading model breaks down. We use their method to estimate the shape of non-shadow regions and, by using robust fitting, use their surface normal model for model-based surface completion. In this section we briefly introduce the approach.

2.1 Statistical Model

A statistical model is trained using a sample of facial surfaces, represented as fields of surface normals (needle-maps) orthographically projected onto the view-plane. In practice these surface normals are derived from a database of laser range scanned faces. Let $\mathbf{n}_p^k \in \mathbb{R}^3$ be the surface normal at the pixel indexed p in the kth training sample. If there are K such samples, then for each pixel the training data provides a distribution of K unit vectors from which we may obtain the local mean direction: $\hat{\mathbf{n}}_p = \frac{\bar{\mathbf{n}}_p}{||\bar{\mathbf{n}}_p||}$ where $\bar{\mathbf{n}}_p = \frac{1}{K}\sum_{k=1}^{K}\mathbf{n}_p^k$.

Principal components analysis (PCA) is used to derive the principal modes of variation of the training sample of facial needle maps. However, this process is complicated by the non-linear nature of unit vectors, namely that a linear combination of unit vectors does not result in a unit vector. For this reason the azimuthal equidistant projection is used to transform distributions of unit vectors to distributions of points on a plane on which a linear analysis can be performed.

We may transform the unit normal $\mathbf{n}_p^k$ to a point $v_p^k \in \mathbb{R}^2$ on the tangent plane to the unit sphere at the point corresponding to the local mean direction. To do so, we use the azimuthal equidistant projection [8]: $v_p^k = \mathrm{AEP}\left(\mathbf{n}_p^k, \hat{\mathbf{n}}_p\right)$, where the second argument is the centre of projection. The resulting distribution of points on the plane retain their variance with respect to the mean and a standard linear PCA can be applied.

A field of surface normals of dimension N may be represented by the long vector: $\mathbf{U}^k = [v_p^k, \ldots, v_N^k]^T$. To perform PCA we form the data matrix: $\mathbf{D} = [\mathbf{U}^1|\ldots|\mathbf{U}^K]$ and find the eigenvectors and eigenvalues of the covariance matrix: $\mathbf{L} = \frac{1}{K}\mathbf{D}\mathbf{D}^T$. We denote the eigenvector with the ith largest eigenvalue $\mathbf{e}_i$ and write the matrix of eigenvectors: $\mathbf{P} = (\mathbf{e}_1|\ldots|\mathbf{e}_K)$. A field of surface normals $\mathbf{U}$ may be projected onto the span of the eigenvectors and represented using a vector of parameters of length K:

$$\mathbf{b} = \mathbf{P}^T\mathbf{U} \tag{1}$$

2.2 Shape-from-shading

If I_p is the measured image brightness at pixel location p, then assuming Lambertian reflectance and unit albedo: $I_p = \mathbf{n}_p.\mathbf{s}$, where $\mathbf{s}$ is the light source direction. Self-shadowing may be taken into account in this formulation by setting negative values to zero: $I_p = \max(\mathbf{n}_p.\mathbf{s}, 0)$. Shape-from-shading aims to invert this equation to recover the local normal estimate from the intensity value. In general, the surface normal can not be recovered from a single brightness measurement since the normal has two degrees of freedom. However, the image irradiance equation does constrain the angle between the light source and surface normal: $\theta_p = \arccos(I_p)$. In the Worthington and Hancock [9] framework, this constraint is satisfied as a hard constraint by forcing the normal $\mathbf{n}_p$ to lie on a cone whose axis is the light source direction and whose apex angle is θ_p.

Smith and Hancock [7] combine this local irradiance constraint with the global constraint provided by the statistical surface normal model in an iterative manner. The method iterates between enforcing compliance with the image irradiance constraint and forcing the field of normals to lie within the span of the eigenvectors of the statistical model. The method commences from an initial estimate of the normals and the field of normals at iteration (t) is: $\mathbf{U}^{(t)}$. The vector of model parameters $\mathbf{b}^{(t)}$ describing the best fit to this

estimate is given by (1). The corresponding field of normals (which may not satisfy the image irradiance constraint) is given by: $\mathbf{U}'^{(t)} = \mathbf{P}\mathbf{b}$. By rotating a normal $\mathbf{n}_p'^{(t)}$ from this field to the closest position on its local irradiance cone, compliance with the image irradiance equation is ensured. This gives the estimate of the normals for the next iteration:

$$\mathbf{n}_p^{(t+1)} = \Theta_p \mathbf{n}_p'^{(t)} \tag{2}$$

where Θ_p is a rotation matrix which restores the normal to its cone.

The algorithm can be summarised as follows:

1. Initialise normals using local mean direction: $\mathbf{n}_p^{(0)} = \Theta_p \hat{\mathbf{n}}_p$, set iteration $t = 1$.
2. Each normal undergoes azimuthal equidistant projection and is stacked to form long vector: $\mathbf{U}^{(t)}$.
3. Estimate parameter vector: $\mathbf{b}^{(t)} = \mathbf{P}^T \mathbf{U}^{(t)}$.
4. Off-cone field of normals $\mathbf{n}_p'^{(t)}$ given by inverse azimuthal equidistant projection of: $\mathbf{U}'^{(t)} = \mathbf{P}\mathbf{b}^{(t)}$.
5. Rotate normals to restore irradiance constraint: $\mathbf{n}_p^{(t+1)} = \Theta_p \mathbf{n}_p'^{(t)}$.
6. Set t=t+1 and iterate to 2.

3 Robust Fitting

The algorithm described above converges quickly and provides a good estimate of facial shape from a single image when the illumination is close-to-frontal. Moreover, enforcing compliance with the image irradiance equation (2) and projection onto the statistical model (1) are operations that can be implemented efficiently as matrix multiplications. However, the local shading model ignores the effect of cast shadows. In these regions, the low intensity will be erroneously interpreted as having a normal whose angle with the light source direction is large. When the field of normals is projected onto the statistical model, the parameter vector will be disrupted by these erroneous estimates. The result is that when shadow regions become significant, the algorithm 'walks away' from the true solution.

We propose replacing (1) with a robust fit, which aims to fit the model to non-shadow regions only. Regions for which we have low confidence in the normal estimates are assigned a low weight in the fit. We use two ingredients to calculate this weighting. First, the current estimate of the shape is used to calculate a cast shadow map. Regions which lie inside cast shadows are assigned zero weight. Second, we use the residual at each point to calculate a weight based on the Huber [6] M-estimator. We initialise the process by using the cast shadow map calculated for the mean face and residuals which measure the distance between the estimated and local mean normal directions.

3.1 Calculating Residuals

Assume the current robust estimate of the best-fit parameter vector is $\mathbf{b}^{(t)}$. The residual η_p at point p is the angular distance required to restore the normal $\mathbf{n}_p'^{(t)}$ given by the parameter vector to its cone. This is given by:

$$\eta_p = \| \theta_p - \left[\mathbf{n}_p'^{(t)} . \mathbf{s} \right] \| \tag{3}$$

3.2 Shadow map estimation

Using a standard surface integration algorithm [4] we can estimate the surface height function $z_p^{(t)}$ from the best-fit field of surface normals $\mathbf{n}_p'^{(t)}$. From the estimated surface height function and light source vector we can estimate a binary cast shadow map at iteration (t) using the function:

$$\text{shadow}(p, z^{(t)}, \mathbf{s}) = \begin{cases} 0 & \text{if pixel } p \text{ is in cast shadow} \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

This function is implemented using a simple ray-tracer.

3.3 Weighting

We combine these two ingredients to assign a weight to each pixel p as follows:

$$w_\sigma(\eta_p) = \begin{cases} 0 & \text{if shadow}(p, z, \mathbf{s}) = 0 \\ 1 & \text{if shadow}(p, z, \mathbf{s}) = 1 \wedge |\eta_p| < \sigma \\ \frac{\sigma}{|\eta_p|} & \text{otherwise} \end{cases} \tag{5}$$

For regions lying inside the estimated cast shadows, a weight of zero is assigned. In the remaining regions, the Huber weight is calculated using the residual. The width parameter σ controls how much of the data is treated as outliers. We use a robust estimate of the standard deviation of the data for this parameter, calculated from the median absolute deviation.

3.4 Weighted Fit

We construct a diagonal matrix of weights: $\mathbf{W} = \text{diag}\left(w_\sigma(\eta_1), \ldots, w_\sigma(\eta_N)\right)$. We use the weights to robustly estimate the parameter vector:

$$\mathbf{b}^{(t)} = C\mathbf{P}^T\mathbf{W}\mathbf{U}^{(t)} \tag{6}$$

where C is a constant which compensates for the overall scaling effect of $\mathbf{W}$ on $\mathbf{b}$. If C is set to the sum of the reciprocals of the weights, this amounts to a one-step weighted least squares fit. However, this approach becomes unstable when a large number of normals have been assigned low weights. The result is over-fitting to sparse and potentially noisy data. We overcome this problem by introducing a control parameter, ε, which represents the trade-off between goodness of fit and distance from the mean. Accordingly, we set C as follows:

$$C = \varepsilon \sum_{p=1}^{N} \frac{1}{w_\sigma(\eta_p)} \tag{7}$$

where ε is allowed to lie in the interval $[0, 1]$. If $\varepsilon = 0$, the result at every iteration is the mean field of normals since the parameter vector $\mathbf{b}^{(t)}$ will be zero. If $\varepsilon = 1$, a one-step weighted least squares fit is performed. For robust performance on real world data, a value somewhere in between is preferable. We use our proposed weighted fit (6) to replace step 3 of the algorithm given in Section 2.2. The weights are recalculated at each iteration using residuals calculated from the previous iteration.

3.5 Surface Completion

On convergence, the normal $\mathbf{n}_p^{(\text{final})}$ satisfies the image irradiance constraint. However, pixels whose corresponding weight is low are considered likely to lie in a cast shadow. In this case, enforcing data-closeness will result in a poor estimate of the normal direction. In these regions we in-fill the surface using the robust fit of the statistical model. Hence, we take the weighted average, $\mathbf{n}_p^{\text{combined}}$, of the best-fit and on-cone normal at each pixel, using $w_\sigma(\eta_p)$ as the weight for $\mathbf{n}_p^{(\text{final})}$ and $1 - w_\sigma(\eta_p)$ as the weight for $\mathbf{n}_p'^{(\text{final})}$.

4 Experiments

In this section we evaluate the performance of the method for shadow removal, albedo estimation and facial shape reconstruction. We train our statistical model on a sample of 100 facial needle-maps. The data is acquired from the 3DFS dataset [1] which consists of 100 high resolution scans of subjects in a neutral expression. The scans were collected using a *Cyberware*$^{\text{TM}}$ 3030PS laser scanner. The database is pre-aligned, registration being performed using the optical flow correspondence algorithm of Blanz and Vetter [3]. For ground truth, we use a leave-one-out strategy in which we train the model with 99 sets of data, leaving the remaining needle-map as out-of-sample ground truth.

We begin by applying the method to known ground truth data allowing us to quantitatively asses the performance of the approach. We then apply the method to real world images, demonstrating the robustness of the approach under real world conditions. For the real world images, we show reconstructions and reilluminations of images from the Yale-B database [5]. These contain albedo variation and cast shadows.

4.1 Ground Truth Data

In Fig. 2 we demonstrate the performance of our method on ground truth data. We apply our algorithm to a selection of images of rendered ground truth needle-maps including cast shadows. In column 1 we show the input images. The needle-maps of the out-of-sample subjects are rendered with Lambertian reflectance and a point light source with direction $\mathbf{s} = (-1, 0, 1)$, i.e. 45° from the viewing direction. We also simulate the effect of cast shadows using the shadow map shown in column 2. shadow$(p, z, \mathbf{s})$ is calculated from ground truth depth data. In column 3 we show the weight function $w_\sigma(\eta_p)$ for each pixel. It is clear that regions in cast shadow have been successfully down-weighted. In column 4 we show the needle-map $\mathbf{n}^{\text{combined}}$ calculated from the input image, rendered with frontal illumination. For comparison, in column 5 we show the ground truth needle-map similarly illuminated. There is a good agreement between the two, even in areas in which no information was present in the input image (i.e. those in cast shadows). This suggests that the robust fit of the model has recovered globally accurate shape information, and has filled-in the shadowed areas of the face. The mean surface normal error was typically $< 8^\circ$ across the whole needle-map. Finally in column 6 we show the shadow map shadow$(p, z^{\text{combined}}, \mathbf{s})$, where z^{combined} is the height map integrated from $\mathbf{n}^{\text{combined}}$. Again, there is a good agreement between columns 2 and 6, suggesting that this represents a viable means to estimate regions which are in cast shadow.

In Fig. 3(a) we examine the influence of illumination direction on the accuracy of the recovered surface normals. Once again we render ground truth, out-of-sample needle-

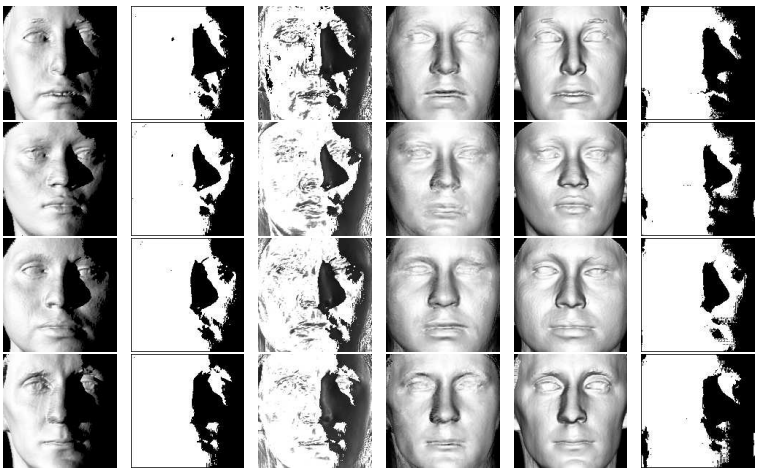


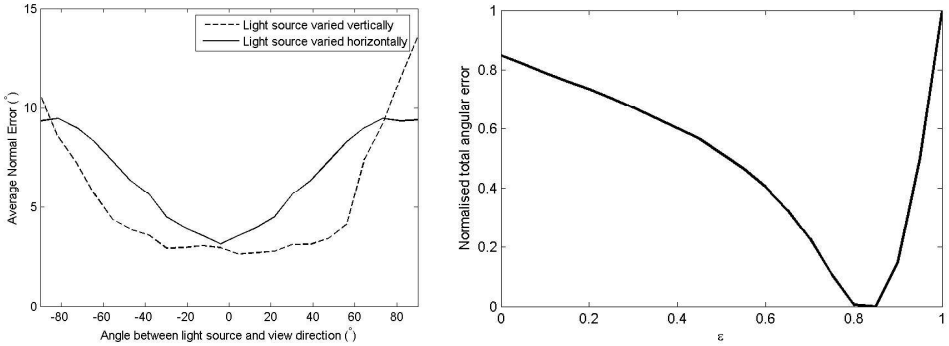
Figure 2: Fitting to images of ground truth needle-maps rendered with Lambertian reflectance and cast shadows.

maps with Lambertian reflectance and a point light source for which we simulate cast shadows. The direction of the light source is varied through a horizontal and vertical arc, i.e. from left to right and from top to bottom. Fig. 3(a) plots the average surface normal error against the angle between the light source and viewing direction. The arc from left to right ($\mathbf{s} = (-1,0,0),\ldots,(1,0,0)$) is shown as a solid line, while the arc from top to bottom ($\mathbf{s} = (0,1,0),\ldots,(0,-1,0)$) is shown as a broken line. The plot demonstrates that our method recovers globally accurate shape information, even when the lighting direction is extreme and hence, much of the face is in shadow. For example, illumination from the extreme right or left still results in an average normal error of less than 10°.

4.2 Real World Data

In Figure 3(b) we begin by providing a quantitative analysis of the control parameter ε. We use our method to recover facial shape from an image which contains significant cast shadows and for which ground truth shape information is known. We show the (normalised) total angular error between the estimated and actual surface normals as ε is varied. It is clear there is a minima at approximately $\varepsilon = 0.8$ and that any greater value sharply increases the error due to over-fitting. As ε tends to zero there is a smoother degradation as the recovered shape tends towards the mean face. For real world data, a more conservative setting of ε is required, but its effect is similar. In Figure 4 we show the shape recovered from a real world image as ε is varied. The effects of overfitting are clear as ε tends to one.

In Figure 5 we show results of applying our method to real world images. In the first column we show input images in which the subjects are illuminated by a light source 50° to the left. In the second column we show the recovered shape, rendered with constant albedo and frontal illumination. It is clear that the surface in the shadowed regions has been convincingly filled in. Finally, in the third column we show the shadow map calcu-



(a) (b)

Figure 3: (a) Average angular error of the recovered normal versus illumination direction (b) Plot of parameter ε versus normalised total angular error

$\varepsilon = 0.4$ $\varepsilon = 0.6$ $\varepsilon = 0.8$ $\varepsilon = 1$

Figure 4: Effect of ε parameter fitting to real-world image

lated from the recovered shape. These agree well with the cast shadows visible in the first column. In particular, note the bump present in the shadow cast by the nose in the third row.

In Fig. 6, we demonstrate the method of real images which contain cast shadows as well as albedo variations. The first row shows the input images of a single subject under varying illumination. The subject is a challenging choice due to the large albedo variations caused by facial hair. The light source is moved in an arc along the horizontal axis to subtend an angle of -50°, -25°, 0°, 25° and 50° with the viewing direction. We use our method to estimate the normals, albedo and shadow map. We use facial symmetry to fill-in the missing albedo values for the shadow regions. In the second row we show the estimated cast shadow map. Here, the cast shadows caused by the nose seem to correspond well with the input images. Finally in the third row, we show the recovered needle-maps rendered with the estimated albedo and frontal lighting, effectively correcting for variation in input lighting. These synthesised images are of a good quality, even under large changes in illumination and manage to remove much of the effect of the cast shadows.

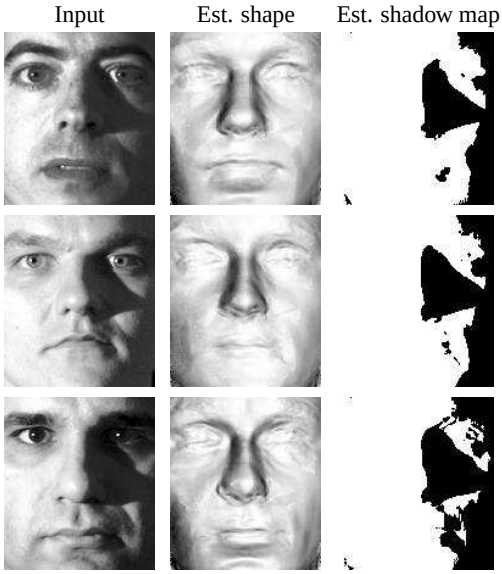


Figure 5: Results for real world images.

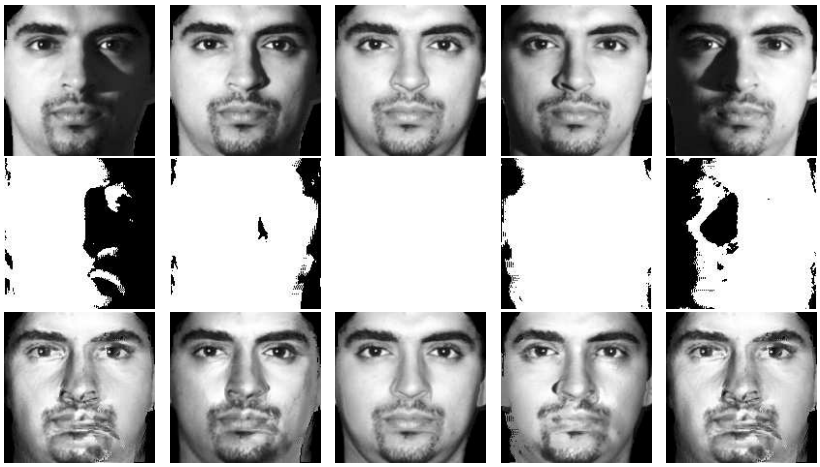


Figure 6: Novel viewpoint of the surface recovered from the input image in the top row rotated 45° about the vertical axis.

5 Conclusions

We have presented an approach to the problem of estimating cast shadows from single face images. We simultaneously estimate the facial shape using shape-from-shading in non-shadow regions and class-based surface completion in shadowed regions. Our approach yields a surface which accurately reconstructs the cast shadows visible in the image, whilst realistically filling in the missing surface in cast shadow regions. We are able to recover accurate facial shape in the presence of large cast shadows. The method may prove useful for cast shadow removal as a preprocessing step to further facial analysis. In future work we will investigate exploiting the shape cue provided by the contour of the cast shadow, which may serve to aid 3D reconstruction by providing information about the profile of the nose.

References

[1] USF HumanID 3D Face Database, Courtesy of Sudeep. Sarkar, University of South Florida, Tampa, FL.

[2] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. PAMI*, 25(2):218–233, 2003.

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. PAMI*, 25(9):1063–1074, 2003.

[4] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. PAMI*, 10(4):439–451, 1988.

[5] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. PAMI*, 23(6):643–660, 2001.

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

[7] W.A.P. 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.

[8] J. P. Snyder. *Map Projections–A Working Manual, U.S.G.S. Professional Paper 1395*. United States Government Printing Office, Washington D.C., 1987.

[9] P. L. Worthington and E. R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. PAMI*, 21(12):1250–1267, 1999.

[10] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Trans. PAMI*, 21(8):690–706, 1999.