

Facial Shadow Removal

William A.P. Smith and Edwin R. Hancock

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

Abstract. In this paper we demonstrate how to recover surface shape from single images of faces using shape-from-shading when shadows are present. We make use of a statistical representation of the distribution of surface normal directions based on the equidistant azimuthal projection. This allows us to develop a statistical model of the variations in facial shape in the surface normal domain. We show how ideas from robust statistics can be used to fit the model to facial images in which there is significant self-shadowing. The method is evaluated on both synthetic and real-world images. It is demonstrated to effectively fill-in the facial surface when more than 30% of the area is subject to self-shadowing.

1 Introduction

The problem of reconstructing the surface height function of a face from a single image is a challenging one that has eluded efforts in shape-from-shading for several decades [1,2]. The goal is to use the image irradiance equation to recover estimates of local surface orientation, and then to recover surface height by integrating the field of surface normals. Unfortunately, there are a number of well documented problems that frustrate this task [1]. The first of these is that the recovered field of surface normals is subject to concave-convex shape ambiguities, and this can lead to the implosion of facial features. This effect can lead to facial features becoming inverted (such as the nose), and the exaggeration of the relief of others (for instance the cheeks). The second problem is that of self shadowing. This occurs when the light source is at an oblique angle to the face, and the nose casts a shadow and the eye-sockets are also in shadow.

It is for these reasons that the problem of facial shape-from-shading is addressed either using domain specific constraints or using a face-specific shape-model. For instance Zhao and Chellappa [1] have shown how improved shape recovery can be achieved using a symmetric shape-from-shading method. Atick et al [3] have used a model based method constructed from range-data. Smith and Hancock [4] have developed a statistical model of surface orientation distribution trained on gradient data from range images. The model is fitted to intensity data using geometric constraints on the surface normal direction provided by Lambert's law. In related work, Vetter and Blanz have shown how recognition can be performed by fitting a surface model to brightness data, and how the fitted model can be used for realistic view synthesis [2]. Finally, it is

worth noting that if multiple images from a fixed viewing direction and variable light source direction are available, then photometric stereo can be used for accurate facial surface recovery [5].

Although the methods listed above lead to realistic shape recovery, they do not work well when there is significant self shadowing. Our aim in this paper is therefore to overcome the problem of self-shadowing by fitting a statistical model to image brightness data using robust statistics. The model has been described in our previous work. It captures variations in facial shape using ideas from cartography. Here the distribution of surface normal directions at different locations on the face is captured on a unit sphere. We convert this distribution of surface normal directions into a distribution of points using the azimuthal equidistant projection. The modes of facial variation are captured by applying principal components analysis to the point-distributions. We demonstrate how the statistical model may be fitted to image brightness data using robust statistics, so as to satisfy constraints provided by Lambert's law. Here the robust statistics treat the shadow regions as outliers. We use the difference between measured and fitted brightness values to compute a shadow-weight. The weights are used to exclude shadow regions in the shape-parameter estimation process. The parameter update scheme is based on M-estimators.

We experiment with the resulting shape-recovery method on both synthetic images with known ground truth and real-world images from the Yale-B database. The results indicate that the method works well even when the angle between the light source and the image-normal exceeds 60 degrees. Moreover, the fitting method is able to reliably fill-in shadowed regions of the face.

2 A Statistical Surface Normal Model

Constructing a statistical model that captures the statistical distribution of directional data is not a straightforward task. To overcome the problem, we draw on ideas from cartography. Our starting point is the *azimuthal equidistant* projection [6]. This projection has the important property that it preserves the distances between locations on the sphere. Another useful property of this projection is that straight lines on the projected plane through the centre of projection correspond to great circles on the sphere. We exploit these properties to generate a local representation of the field of surface normals. We commence with a set of needle-maps, i.e. fields of surface normals which in practice are obtained either from range images or shape-from-shading. We begin by computing the mean field of surface normals. The surface normals are represented using elevation and azimuth angles on a unit sphere. At each image location the mean-surface normal defines a reference direction which we use to construct an azimuthal equidistant projection for the distribution of surface normals at this point. The distribution of points on the projection plane preserves the distances of the surfaces normals on the unit sphere with respect to the mean surface normal, or reference direction. We then construct a deformable model over the set of surface normals by applying the Cootes and Taylor [7] point distribution model

to the co-ordinates that result from transforming the surface normals from the unit sphere to the tangent plane under azimuthal equidistant projection. More details of this model are given in [4].

3 Fitting the Model to an Image

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

If $I(i, j)$ is the measured image brightness at location (i, j) , then $I(i, j) = \omega(i, j) [\mathbf{n}(i, j) \cdot \mathbf{s}]$ according to Lambert's law, where \mathbf{s} is the light source direction and ω is the albedo. We begin by assuming constant and unit albedo (the Lambertian remapping process [8] normalises the brightest point to unity) and return to this later. In general, the surface normal \mathbf{n} can not be recovered from a single brightness measurement since it has two degrees of freedom corresponding to the elevation and azimuth angles on the unit sphere. In the Worthington and Hancock [9] iterative shape-from-shading framework, data-closeness is ensured by constraining the recovered surface normal to lie on the reflectance cone whose axis is aligned with the light-source vector \mathbf{s} and whose opening angle is $\alpha = \arccos I$. At each iteration the surface normal is free to move to an off-cone position subject to smoothness or curvature consistency constraints. However, the hard irradiance constraint is re-imposed by rotating each surface normal back to its closest on-cone position. This process ensures that the recovered field of normals satisfies the image irradiance equation after every iteration. The framework is initialised by placing the surface normals on their reflectance cones in the direction opposite to that of the local image gradient.

Our approach to fitting the model to intensity images uses the fields of surface normals estimated using the geometric shape-from-shading method described above. This is an iterative process in which we interleave the process of fitting the statistical model to the current field of estimated surface normals, and then re-enforcing the data-closeness constraint provided by Lambert's law by mapping the surface normals back onto their reflectance cones. If the data is of dimensions $M \times N$, the surface normal model is encapsulated in the $2MN \times K$ matrix $\mathbf{P} = (\mathbf{e}_1 | \mathbf{e}_2 | \dots | \mathbf{e}_K)$ formed from the leading K principal eigenvectors of the covariance matrix of the training samples under azimuthal equidistant projection. In other words, they are the modes of variation of the statistical surface normal model. The fitting algorithm can therefore be summarised as follows:

1. Initialise the field of surface normals \mathbf{n} .
2. Each normal in the estimated field \mathbf{n} undergoes an azimuthal equidistant projection to give a vector of transformed coordinates \mathbf{v} .
3. The vector of best fit model parameters is $\mathbf{b} = \mathbf{P}^T \mathbf{v}$.
4. The corresponding vector of transformed coordinates is $\mathbf{v}' = \mathbf{P} \mathbf{P}^T \mathbf{v}$.

5. Using the inverse azimuthal equidistant projection find \mathbf{n}' from \mathbf{v}' .
6. Find \mathbf{n}'' by rotating each normal in \mathbf{n}' back to their closest on-cone position.
7. Stop if the difference between \mathbf{n} and \mathbf{n}'' indicates convergence.
8. Make $\mathbf{n} = \mathbf{n}''$ and return to step 2.

The method hence combines a strict global constraint (projection onto the statistical model) with a hard local constraint (satisfaction of the image irradiance equation). We find the algorithm converges rapidly and offers stable performance on real world images. However, a number of obstacles are encountered when this simple approach is applied to real images. The simple reflectance model given Lambert's law assumes constant albedo and ignores the effect of cast shadows (regions in which the light source is intercepted by another part of the surface). The result is that the fit to the statistical model is subject to a systematic error and becomes increasingly inaccurate when regions of low albedo dominate (for instance in the presence of facial hair) or when cast shadows become significant (as the light source direction is more extreme). In this case, a significant portion of the face may be in shadow and fitting the statistical model globally results in erroneous shape parameter estimates.

4 Robust Statistics

It is therefore clear that we require a more robust means to fit our statistical model to a potentially noisy observed field of normals, \mathbf{n} . The quality of this fit can be measured by calculating the distance between the observed and fitted normals on the projection plane. If \mathbf{b} is the estimated parameter vector, the vector of residuals is given by $\mathbf{R} = \|\mathbf{v} - (\mathbf{P}\mathbf{b})\|$ and the residual at point p is $\eta_p = \sqrt{R(2p-1)^2 + R(2p)^2}$. The standard least squares fit given above, minimises the quantity:

$$\mathbf{b}^* = \arg \min_{\mathbf{b}} \sum_{i=1}^N \eta_i^2.$$

This approach is unstable in the presence of outlying data, such as normals erroneously estimated from regions of low albedo or in cast shadow. In particular, the effect of outliers is to severely distort the estimated facial shape.

In this paper, we turn to the apparatus of robust statistics to help overcome this problem. M-estimators (maximum likelihood type estimators) aim to reduce the effect of outliers by replacing the squared residuals η_i^2 by a kernel function that limits the effects large residuals:

$$\mathbf{b}^* = \arg \min_{\mathbf{b}} \sum_{i=1}^N \rho_{\sigma}(\eta_i) \quad (1)$$

where ρ is a robust kernel with width parameter σ .

The influence of a residual on the parameter estimate under a given M-estimator can be studied by examining its *influence function*, ψ_{σ} . This is the

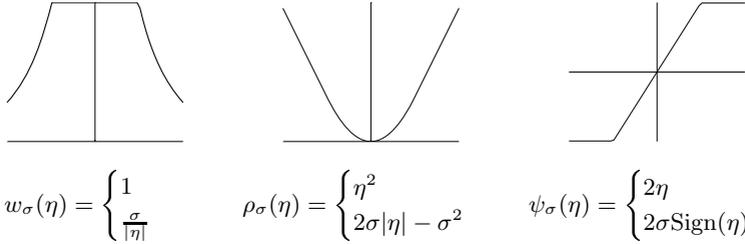


Fig. 1. Huber’s M-estimator

derivative of the error kernel: $\psi_\sigma(\eta_p) = \frac{\partial \rho_\sigma(\eta_p)}{\partial \eta_p}$. Hence, in the least squares case where $\rho_\sigma(\eta) = \eta^2$, the influence of a datum is $\psi_\sigma(\eta) = 2\eta$ and therefore increases linearly with the size the error. This is the source of the lack of robustness in least-squares estimation.

We propose a robust solution to (1) using a simple one-step weighted least squares approximation. To do so, we make use of the *weight function*, w_σ , which is related to the influence function by: $w_\sigma(\eta_p) = \frac{\psi_\sigma(\eta_p)}{\eta_p}$. The standard least-squares estimator applies a constant weight to each datum. On the other hand, an error kernel such as Huber’s estimator [10] down-weights a datum once its residual exceeds σ :

$$\rho_\sigma(\eta) = \begin{cases} \eta^2 & \text{if } |\eta| < \sigma \\ 2\sigma|\eta| - \sigma^2 & \text{otherwise} \end{cases} \quad w_\sigma(\eta) = \begin{cases} 1 & \text{if } |\eta| < \sigma \\ \frac{\sigma}{|\eta|} & \text{otherwise} \end{cases} \quad (2)$$

We show the weight function, error kernel and influence function for Huber’s M-estimator in Fig 1. This is the M-estimator we use in the remainder of this paper.

We can incorporate the Huber weights into the least squares fit by constructing a diagonal matrix of weights: $\mathbf{W} = \text{diag}(w_\sigma(\eta_1), \dots, w_\sigma(\eta_N))$. Our one-step weighted least squares approximation of \mathbf{b} is given by:

$$\mathbf{b}^{(t)} = C\mathbf{P}^T \mathbf{W}\mathbf{v}^{(t)} \quad (3)$$

where C is a constant which compensates for the overall scaling effect of \mathbf{W} on \mathbf{b} : $C = N\text{Tr}(\mathbf{W}^{-1})$.

Computing (3) requires an initial estimate of \mathbf{W} and hence the residuals. We therefore commence by calculating a least squares fit from which we can calculate the residuals η_p . This is equivalent to calculating (3) with an identity weight matrix, i.e. $\mathbf{W} = \mathbf{I}_N$. For a subsequent iteration t of the algorithm, we can use the weights calculated from the residuals at iteration $(t - 1)$.

Implicit in the discussion above is that we have a means to estimate the standard deviation of the residual errors σ , which acts as the width parameter of the function ρ . A robust estimate of σ is required in order to distinguish outliers from inliers. For this reason, we use the *median absolute deviation* (MAD):

$$\text{MAD} = \text{median}(|\eta_p - \text{median}(\eta_p)|), p = 1 \dots N$$

which is related to the standard deviation by $\sigma = 1.4826 \times \text{MAD}$.

4.1 Combining and Classifying

Upon convergence, if a pixel p has weight $w_p^{final} \approx 1$, this indicates a high confidence that the on-cone normal \mathbf{n}_p'' is reliable. However, as the weight tends to 0, the on-cone normal \mathbf{n}_p'' is likely to be erroneous due to violation of the assumptions of Lambert’s law, e.g. non-constant albedo or lying in a cast shadow region. In this case, a more accurate estimate of \mathbf{n}_p^{final} is given by the robust fit of the model to the global field of normals.

For this reason, our best estimate of the underlying shape of the face is a weighted combination, in which pixels with a low weight are given a higher proportion of the normal from the model fit \mathbf{n}_p' and a lower proportion of the on-cone normal \mathbf{n}_p'' , vice-versa for pixels with a high weight. This gives: $\mathbf{n}_p^{combined} = \frac{\bar{\mathbf{n}}_p}{\|\bar{\mathbf{n}}_p\|}$, where $\bar{\mathbf{n}}_p = w_p^{final} \mathbf{n}_p'' + (1 - w_p^{final}) \mathbf{n}_p'$.

With the estimated facial shape to hand, we may now go further and distinguish between pixels of low albedo and those in cast shadow regions. We may recover the surface height z_p by applying a standard surface integration method [11] on the field of normals $\mathbf{n}^{combined}$. Using a simple ray-tracing algorithm, we can assign a binary cast-shadow map:

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

For non-shadow regions, the albedo ω_p can be estimated by rearranging the image irradiance equation: $\omega_p = \frac{I_p}{\mathbf{n}_p^{combined} \cdot \mathbf{s}}$.

5 Experimental Results

In this section we present experiments using the technique described above. We begin by applying the method to known ground truth data allowing us to quantitatively assess the performance of the approach. We then apply the method to real world images, demonstrating the robustness of the approach under real world conditions.

We train our statistical model on a sample of 100 facial needle-maps. The data is acquired from the 3DFS dataset [12] which consists of 100 high resolution scans of subjects in a neutral expression. The scans were collected using a *Cyberware*TM 3030PS laser scanner. The database is pre-aligned, registration being performed using the optical flow correspondence algorithm of Blanz and Vetter [2]. 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.

For real world images, we show reconstructions and reilluminations of images from the Yale-B database [5]. These contain albedo variation and cast shadows.

5.1 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

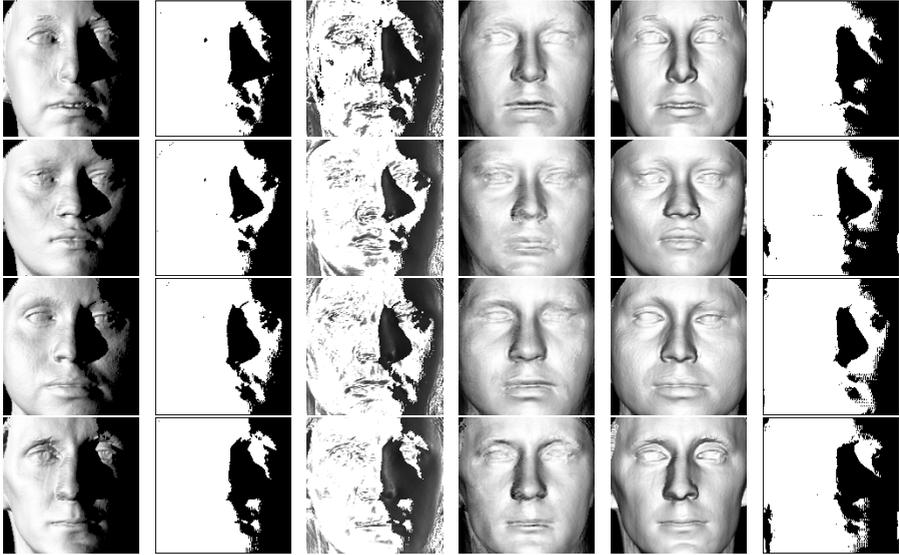


Fig. 2. Fitting to images of ground truth needle-maps rendered with Lambertian reflectance and cast shadows

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. $\text{shadow}(z, \mathbf{s})$ 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 $\text{shadow}(z^{\text{combined}}, \mathbf{s})$, 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.

5.2 Real World Data

In Fig. 3, we demonstrate the quality of the shape information our method can recover from real world images. From the input images in the top row, we use our method to estimate the needle-map $\mathbf{n}^{\text{combined}}$ and use the surface recovery



Fig. 3. Novel viewpoint of the surface recovered from the input image in the top row rotated 45° about the vertical axis

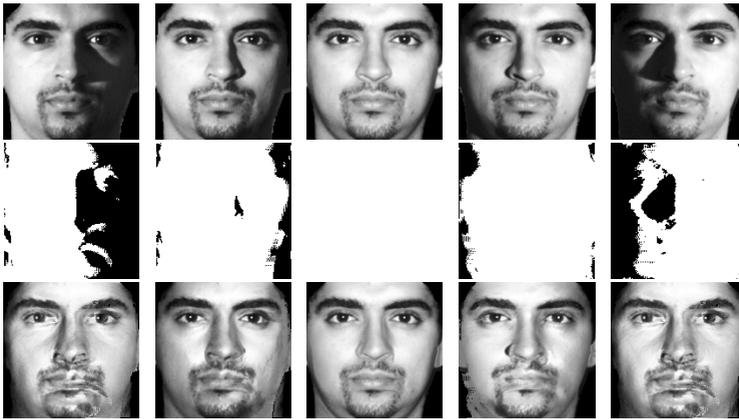


Fig. 4. Novel viewpoint of the surface recovered from the input image in the top row rotated 45° about the vertical axis

method of Frankot and Chellappa [11] to integrate the normals into a surface. We show these surfaces rendered with the estimated albedo and Lambertian reflectance, rotated 45° about the vertical axis. The images show considerable stability under large change in viewpoint.

In Fig. 4, we demonstrate the method on 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.

6 Conclusions

We have shown how a statistical model of facial shape, couched in terms of distributions of surface normal directions, can be fitted to images of shadowed faces using robust statistics. We fit the statistical model globally, but use robust statistics to ensure that regions of low albedo or which fall into a cast shadow have little or no impact on the parameter estimate. The technique is capable of recovering a useful estimate of facial shape, even when significant portions of the face are entirely in shadow.

References

1. Zhao, W.Y., Chellappa, R.: Illumination-insensitive face recognition using symmetric SFS. In: Proc. CVPR. (2000) 286–293
2. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. *IEEE Trans. PAMI* **25** (2003) 1063–1074
3. Atick, J.J., Griffin, P.A., Redlich, A.N.: Statistical approach to SFS: Reconstruction of 3D face surfaces from single 2D images. *Neural Comp.* **8** (1996) 1321–1340
4. Smith, W., Hancock, E.R.: Recovering facial shape and albedo using a statistical model of surface normal direction. In: Proc. ICCV. (2005) 588–595
5. Georgiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. PAMI* **23** (2001) 643–660
6. Snyder, J.P.: *Map Projections—A Working Manual*, U.S.G.S. Professional Paper 1395. United States Government Printing Office, Washington D.C. (1987)
7. Cootes, T.F., Taylor, C., Cooper, D., Graham, J.: Training models of shape from sets of examples. In: Proc. BMVC. (1992) 9–18
8. Smith, W., Robles-Kelly, A., Hancock, E.R.: Reflectance correction for perspiring faces. In: Proc. ICIP. (2004) 1389–1392
9. Worthington, P.L., Hancock, E.R.: New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. PAMI* **21** (1999) 1250–1267
10. Huber, P.: *Robust Statistics*. Wiley, Chichester (1981)
11. Frankot, R.T., Chellappa, R.: A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. PAMI* **10** (1988) 439–451
12. USF HumanID 3D Face Database, Courtesy of Sudeep. Sarkar, University of South Florida, Tampa, FL.