

# Robust Estimation of Albedo for Illumination-invariant Matching and Shape Recovery

Soma Biswas, *Student Member, IEEE*, Gaurav Aggarwal, *Student Member, IEEE*,  
and Rama Chellappa, *Fellow, IEEE*

**Abstract**—We present a non-stationary stochastic filtering framework for the task of albedo estimation from a single image. There are several approaches in the literature for albedo estimation, but few include the errors in estimates of surface normals and light source direction to improve the albedo estimate. The proposed approach effectively utilizes the error statistics of surface normals and illumination direction for robust estimation of albedo, for images illuminated by single and multiple light sources.

The albedo estimate obtained is subsequently used to generate albedo-free normalized images for recovering the shape of an object. Traditional Shape-from-Shading (SFS) approaches often assume constant/piece-wise constant albedo and known light source direction to recover the underlying shape. Using the estimated albedo, the general problem of estimating the shape of an object with varying albedo map and unknown illumination source is reduced to one that can be handled by traditional SFS approaches.

Experimental results are provided to show the effectiveness of the approach and its application to illumination-invariant matching and shape recovery. The estimated albedo maps are compared with the ground truth. The maps are used as illumination-invariant signatures for the task of face recognition across illumination variations. The recognition results obtained compare well with the current state-of-the-art approaches. Impressive shape recovery results are obtained using images downloaded from the web with little control over imaging conditions. The recovered shapes are also used to synthesize novel views under novel illumination conditions.

**Index Terms**—Albedo estimation, shape recovery, image estimation, illumination-invariant matching

## I. INTRODUCTION

A real world object is characterized by its underlying shape and surface properties. These characteristics define the way the object is perceived, irrespective of the view or illumination. Unlike image intensity, these characteristics of an object are invariant to changes in illumination conditions which makes them useful for illumination-invariant matching of objects. Realistic Image based Rendering (IBR) is another application where accurate estimates of shape and albedo (texture) play a very important role. Thus estimating the shape and albedo of an object from an intensity image has been a very important area of research in the field of

Computer Vision and Graphics. Though research on this topic has been underway for over two decades, the difficulty in obtaining accurate estimates and the wide range of applications continue to interest researchers.

### A. Albedo Estimation

Albedo is the fraction of light that a surface point reflects when it is illuminated. It is an intrinsic property that depends on the material properties of the surface. In existing literature, albedo estimation has often been coupled with shape estimation. Given an input image, most methods follow the two-step approach of first recovering the shape of the object and then estimating the surface albedo [1]. A few others simultaneously estimate the shape and albedo of an object [2] [3]. There are also albedo estimation methods whose main goal is shape estimation and albedo is finally incorporated to account for the image reconstruction errors using the estimated shape [4]. Thus, in most approaches, albedo recovery depends on the accuracy of the estimated shape and illumination conditions. Errors in shape and illumination estimates lead to errors in albedo. In this paper, we show how statistical characterization of the errors in estimates of normal and light source directions can be utilized to obtain robust albedo estimates.

The problem of albedo estimation is formulated as an image estimation problem. Given an initial albedo map (obtained using available domain-dependent average shape information), we obtain a robust albedo estimate by modeling the true unknown albedo as a non-stationary mean and non-stationary variance field. Unlike a stationary model, this model can account for the albedo variations present in most real objects. The initial albedo map can be expressed as a sum of the true unknown albedo and a signal-dependent non-stationary noise. The noise term incorporates the errors in surface normal and illumination information. Posing this as an image estimation problem, the albedo is estimated as the Linear Minimum Mean Square Error (LMMSE) estimate of the true albedo.

The theoretical formulation is extended to deal with images illuminated by multiple unknown light sources. We propose an algorithm for estimating the illumination conditions in such scenarios. The algorithm is based on an approximation to the linear subspace property of Lambertian surfaces [5] [6]. The estimated illumination information is used along with a domain-dependent average shape to obtain an initial albedo map. Similar to the single source framework, a robust albedo estimate is obtained by formulating it as an image estimation problem.

Partially funded by a contract from UNISYS Corporation.

S. Biswas and R. Chellappa are with the Center for Automation Research and Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, 20742. Email: {soma, rama}@cfar.umd.edu.

G. Aggarwal was with the Center for Automation Research and Department of Computer Science, University of Maryland, College Park, MD, 20742. He is now with Object Video, Reston, VA. Email: gaggarwal@objectvideo.com.

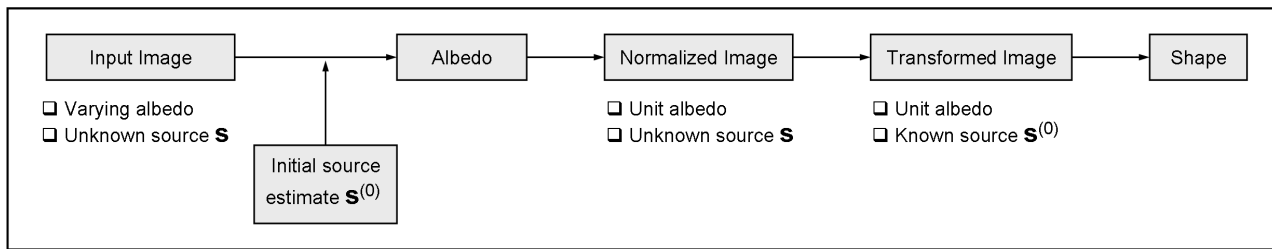


Fig. 1. Schematic diagram of the proposed approach for shape recovery.

### B. Shape Recovery

Like albedo, shape is another intrinsic property of an object which is invariant to changes in pose and illumination conditions. The importance of estimating the shape of an object has probably been the guiding force behind the vast amount of work that has been done to recover shape (shape-from-X) from images or videos. An example in this category is the work done on the problem of Shape from Shading [7] [8]. The goal of SFS research is to recover the 3D shape of an object from a single image. Dealing with an inherently ill-posed problem, SFS approaches typically make simplifying assumptions like constant or piecewise constant albedo and/or known single distant light source. Such assumptions though useful to make the problem more tractable, often limit the applicability of the approaches for real world objects.

We focus on the general SFS problem of estimating the shape and albedo of an object with varying albedo map and unknown illuminant direction from a single image. To this end, we propose an algorithm that transforms the general SFS problem to one of estimating the shape of an object with unit albedo and known illuminant direction that can be addressed using a standard SFS approach.

Starting with initial values (e.g., domain-specific average) for shape and illuminant direction, the algorithm proceeds as follows (Fig. 1)

- 1) **Albedo Estimation and Image Normalization:** Starting with the initial shape and illumination information, the albedo is computed using the image irradiance relation. Naturally, inaccuracies in these initial values lead to errors in the albedo value. We refine the computed albedo using the error statistics of the shape and illumination information in an image estimation framework. The estimated albedo is then used to normalize the input image to obtain an *albedo-free* normalized image. This normalization step reduces the original problem of estimating the shape of an object with varying albedo to one of recovering the shape of an object with unit albedo.
- 2) **Image Transformation:** Traditional SFS approaches usually assume known illuminant direction in addition to constant albedo assumption. In this work, the normalized image is further transformed to a new image of the same object illuminated by the initial estimate of illuminant direction. The transformation is formulated as another estimation problem that utilizes the error statistics of the illuminant direction. This reduces the problem to one of finding the shape of an object with unit albedo and known illuminant direction.
- 3) **Shape Estimation:** Given the transformed image and the

initial illuminant direction, one can use any standard SFS approach to estimate the shape of the object. Using the estimated shape and albedo information, the illuminant direction can also be refined.

A schematic of the proposed approach is shown in Fig. 1. The sequence of steps can be repeated with the new estimates as the initial values for the shape, albedo and illuminant direction for further refinement. A preliminary version of this work has appeared in [9].

### C. Organization of the Paper

The paper is organized as follows. Section II discusses the related works in the literature. The proposed albedo estimation framework is detailed in Section III. This section describes in detail the image estimation framework and the derivation of the required variances. Section IV details the steps involved in shape recovery using the estimated albedo. Experimental results are presented in Section V. Section VI concludes the paper with a brief summary.

## II. PREVIOUS WORK

In this section, we discuss the related works for recovering the surface normals and albedo of an object from an intensity image. Since our approach uses image estimation formulation, we also discuss the related works in image estimation literature.

### A. Related Work on Recovering Surface Normal and Albedo

Recovery of surface normals and albedo of an object has been studied in the computer vision community for a long time. The approaches in the literature can broadly be classified into *SFS-based approaches* and *Model-based approaches*.

**SFS-based Approach:** SFS research [7] [8] aims at recovering the 3D shape of an object from a given image. Estimating the surface normals, albedo and illuminant direction given a single intensity image is inherently ill-posed. In order to make the problem more tractable, SFS approaches often make simplifying assumptions like constant or piecewise constant albedo and known illuminant direction for recovering the shape. But such assumptions often limit the applicability of the approaches for real world objects. Over the years, considerable advances have been made [10] [11] [12] [4] for dealing with objects with varying albedo. Much of the research has been directed towards the use of domain specific constraints to reduce the intractability of the problem for the analysis and estimation of general albedo maps. Often the use of such constraints bridges the gap between pure SFS approaches and the statistical model-based approaches leaving no clear demarcation between the two categories.

Zhao and Chellappa [10] propose an SFS approach to recover shape and albedo for the class of bilaterally symmetric Lambertian objects with a piecewise constant albedo field. Their approach combines the self-ratio image irradiance equation with the standard image irradiance equation to solve for the unknowns. Atick *et al.* [11] reduce the SFS problem to that of parameter estimation in a low-dimensional space using the Principal Component Analysis (PCA) over several hundred laser-scanned 3D heads. Dovgard *et al.* [12] combine the symmetric SFS formulation [10] with the statistical approach to facial shape reconstruction [11] to recover the 3D facial shape from a single image. Smith and Hancock [4] embed a statistical model of facial shape in an SFS formulation. Albedo estimation follows shape estimation to account for the differences between predicted and observed image brightness.

**Model-Based Approach:** Blanz and Vetter [2] propose a 3D morphable model based approach to recognize faces across pose and illumination variations. They represent each face as a linear combination of 3D basis exemplars. Recovery of shape and albedo parameters is formulated as an optimization problem that aims to minimize the difference between the input and the reconstructed image. Romdhani *et al.* [3] provide an efficient and robust algorithm for fitting a 3D morphable model using shape and texture error functions. Their algorithm uses linear equations to recover the shape and texture parameters irrespective of the pose and lighting conditions of the face image. Zhang and Samaras [1] combine spherical harmonics illumination representation [5] [6] with 3D morphable models [2] to recover person-specific basis images. Feature-point based shape recovery is followed by an iterative estimation of albedo and illumination coefficients. Samaras and Metaxas [13] incorporate non-linear holonomic constraints in a deformable model to estimate shape and illuminant direction. Under the assumption of constant albedo, the light source direction and shape are estimated in an iterative manner by fixing one unknown and estimating the other until there is no more change in the illuminant estimate. Zhou *et al.* [14] impose a rank constraint on shape and albedo for the face class to separate the two from illumination using a factorization approach. Integrability and face symmetry constraints are employed to fully recover the class-specific albedos and surface normals. Yue *et al.* [15] use pose-encoded spherical harmonic representation for face recognition and synthesis from a single image. Lee and Moghaddam [16] propose a scheme for albedo estimation and relighting of human faces using a generic 3D face shape. First the average shape is used to determine the dominant light source direction which is then used to obtain an estimate of surface albedo for Lambertian objects. The problem of albedo estimation has also been addressed by lightness algorithms that recover an approximation to surface reflectance in independent wavelength channels [17].

### B. Related Work on Image Estimation Methods

In this paper, albedo estimation is formulated as an image estimation problem. Image estimation being a very mature area in the field of image processing [18], we provide pointers only to a few related papers. The standard Wiener filter is known to be optimal for second order stationary processes [18]. In a stationary model, the statistical properties of the image are globally characterized which makes the stationary Wiener estimation algorithm blur the abrupt changes in the input image. Several

modifications to the standard stationary image model have since been proposed. Hunt *et al.* [19] [20] proposed a non-stationary mean Gaussian image model in which an image is modeled as stationary fluctuations about a non-stationary ensemble mean. Lebedev and Mirkin [21] suggested a composite image model that assumes that an image is composed of many different stationary components, each having a distinct stationary correlation structure. Anderson and Netravali [22] used a subjective error criterion based on human visual system models. The derived non-recursive filter makes a trade-off between the loss of resolution and noise smoothing such that the same amount of subjective noise is suppressed throughout the image. Abramatic and Silverman [23] generalized this procedure and related it to the classical Wiener filter. Naderi and Sawchuk [24] derived a non-stationary discrete Wiener filter for a signal-dependent film-grain noise model which can adapt itself to the local signal statistics given the conditional noise statistics. Kuan *et al.* [25] proposed a non-stationary mean, non-stationary variance image model. A local linear minimum square error filter for images degraded by blur and a class of signal-dependent, uncorrelated noise is derived based on the proposed image model.

### III. ALBEDO ESTIMATION

For Lambertian objects, the diffused component of the surface reflection is modeled using the *Lambert's Cosine Law* given by

$$I = \rho \max(\mathbf{n}^T \mathbf{s}, 0) \quad (1)$$

where  $I$  is the pixel intensity,  $\mathbf{s}$  is the light source direction,  $\rho$  is the surface albedo and  $\mathbf{n}$  is the surface normal of the corresponding surface point. The expression implicitly assumes a single dominant light source placed at infinity. It is worthwhile to note that the Lambert's law in its pure form is non-linear due to the max function which accounts for the formation of attached shadows. Shadow points do not reveal any information about their reflectivity and thus their albedo cannot be estimated from the image.

Let  $\mathbf{n}_{i,j}^{(0)}$  and  $\mathbf{s}^{(0)}$  be the initial values of the surface normal and illuminant direction. The initial surface normal can be the domain dependent average value or any estimate available or obtained from any other method. The Lambertian assumption imposes the following constraint on the initial albedo  $\rho^{(0)}$  obtained

$$\rho_{i,j}^{(0)} = \frac{I_{i,j}}{\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}} \quad (2)$$

where  $\cdot$  is the standard dot product operator. We suppress the explicit max operator by considering only the pixels with positive intensity values. Clearly, more accurate the denominator ( $\mathbf{n}^{(0)} \cdot \mathbf{s}^{(0)}$ ) is, the closer is the obtained initial albedo to its true value  $\rho$ . For most applications, accurate initial estimates of normals and light source direction are not available leading to erroneous  $\rho^{(0)}$ .

Fig. 2 illustrates the nature of errors in the obtained albedo  $\rho^{(0)}$  for a synthetically generated face image using a frontal light source i.e.  $\mathbf{s} = [0, 0, -1]$ . True surface normal information is used for estimating the albedo in this example. One may expect the errors to be larger if inaccurate estimates or average value of surface normals are used. The light source direction is estimated using the method in [26] and the resulting  $\mathbf{s}^{(0)}$  is  $[0.1499, -0.0577, -0.9870]$ . Interestingly, not only is  $\rho^{(0)}$  quite far from the true value for quite a few points, but even the error varies appreciably across pixels. The proposed estimation

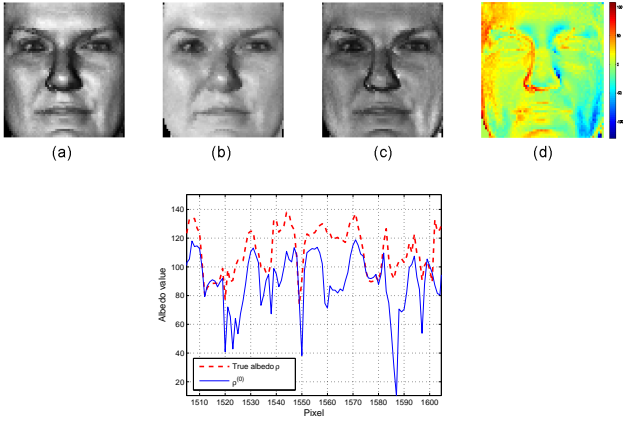


Fig. 2. Illustration of the nature of errors in the initial albedo  $\rho^{(0)}$ . (a) Input Image, (b) True Albedo, (c) Obtained Albedo, (d) Difference Image, (Bottom) Pixel-wise error variation in the initial albedo map  $\rho^{(0)}$ .

framework duly accounts for these variations to obtain a robust albedo estimate.

#### A. Image Estimation Framework

Here we present the image estimation framework to obtain a robust albedo estimate using the initial albedo map which is erroneous due to inaccuracies in surface normal and light source estimates. The expression in (2) can be rewritten as follows

$$\rho_{i,j}^{(0)} = \frac{I_{i,j}}{n_{i,j}^{(0)} \cdot s^{(0)}} = \rho_{i,j} \frac{n_{i,j} \cdot s}{n_{i,j}^{(0)} \cdot s^{(0)}} \quad (3)$$

where  $\rho$ ,  $n$  and  $s$  are the true unknown albedo, normal and illuminant direction respectively.  $\rho^{(0)}$  can further be expressed as follows

$$\rho_{i,j}^{(0)} = \rho_{i,j} + \frac{n_{i,j} \cdot s - n_{i,j}^{(0)} \cdot s^{(0)}}{n_{i,j}^{(0)} \cdot s^{(0)}} \rho_{i,j} \quad (4)$$

Substituting  $w_{i,j} = \frac{n_{i,j} \cdot s - n_{i,j}^{(0)} \cdot s^{(0)}}{n_{i,j}^{(0)} \cdot s^{(0)}} \rho_{i,j}$ , (4) simplifies to

$$\rho_{i,j}^{(0)} = \rho_{i,j} + w_{i,j} \quad (5)$$

This can be identified with the standard image estimation formulation [18]. Here  $\rho$  is the original signal (true albedo), the initial estimate  $\rho^{(0)}$  is the degraded signal and  $w$  is the signal dependent additive noise.

#### B. LMMSE Estimate of Albedo

The Minimum Mean Square Error (MMSE) estimate of the albedo map  $\rho$  given noisy observed map  $\rho^{(0)}$  is the conditional mean

$$\hat{\rho} = E(\rho | \rho^{(0)}) \quad (6)$$

In general, the MMSE estimate is non-linear and depends on the probability density functions of  $\rho$  and  $w$  and is difficult to estimate. Imposing linear constraint on the estimator structure, the LMMSE estimate is given by [27]

$$\hat{\rho} = E(\rho) + C_{\rho\rho^{(0)}} C_{\rho^{(0)}}^{-1} (\rho^{(0)} - E(\rho^{(0)})) \quad (7)$$

Here  $C_{\rho\rho^{(0)}}$  is the cross-covariance matrix of  $\rho$  and  $\rho^{(0)}$ .  $E(\rho^{(0)})$  and  $C_{\rho^{(0)}}$  are the ensemble mean and covariance matrix of  $\rho^{(0)}$  respectively. The LMMSE filter requires the second order statistics of the signal and noise.

The expression for the signal-dependent noise  $w$  can be rewritten as follows

$$w_{i,j} = \frac{(n_{i,j} - n_{i,j}^{(0)}) \cdot s + n_{i,j}^{(0)} \cdot (s - s^{(0)})}{n_{i,j}^{(0)} \cdot s^{(0)}} \rho_{i,j} \quad (8)$$

Assuming the initial values of surface normal  $n^{(0)}$  and light source direction  $s^{(0)}$  to be unbiased, both  $E(w)$  and  $E(w|\rho)$  are zero. Since noise is zero mean,  $E(\rho^{(0)}) = E(\rho)$ . So  $C_{\rho\rho^{(0)}}$  can be written as

$$\begin{aligned} C_{\rho\rho^{(0)}} &= E[(\rho - E(\rho))(\rho^{(0)} - E(\rho^{(0)}))^T] \\ &= C_{\rho} + E[(\rho - E(\rho))w^T] \end{aligned} \quad (9)$$

Similarly, if  $C_w$  is the covariance of the noise term,  $C_{\rho^{(0)}}$  can be written as

$$\begin{aligned} C_{\rho^{(0)}} &= E[(\rho^{(0)} - E(\rho^{(0)}))(\rho^{(0)} - E(\rho^{(0)}))^T] \\ &= C_{\rho} + C_w + E[(\rho - E(\rho))w^T] \\ &\quad + E[w(\rho - E(\rho))^T] \end{aligned} \quad (10)$$

Recalling that  $E(w)$  and  $E(w|\rho)$  are zero,  $E((\rho - E(\rho))w^T) = 0$ . This simplifies (9) and (10) as follows

$$C_{\rho\rho^{(0)}} = C_{\rho} \quad \text{and} \quad C_{\rho^{(0)}} = C_{\rho} + C_w \quad (11)$$

In conventional image estimation problems, the original signal is assumed to be a wide-sense stationary random field. For albedo of real world objects, stationarity may be an oversimplified assumption. Fig. 3(a) shows the albedo of a face. It is evident from the histogram (Fig. 3(b)) which is not Gaussian that the albedo is not a stationary random field.

Here we assume a Non-stationary Mean Non-stationary Variance (NMNV) model [25] for the true albedo  $\rho$ . Unlike the stationary model, the original signal is characterized by a non-stationary mean  $E(\rho)$  which describes the gross structure of the signal. Under this model, the residual component  $(\rho - E(\rho))$  which describes the signal variations is assumed to be a non-stationary white process, i.e. it is statistically uncorrelated and characterized by its non-stationary variance  $\sigma_{i,j}^2(\rho)$ . Fig. 3(c) shows the normalized unit variance residual image. The normalized image seems to have Gaussian-like histogram (Fig. 3(d)) which justifies the NMNV model for the true unknown albedo [25].

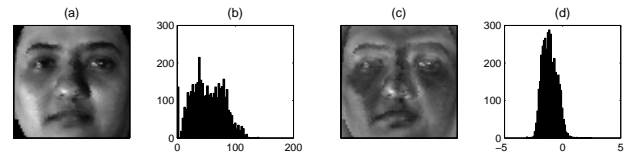


Fig. 3. Non-stationary mean and non-stationary variance model for the true albedo. (a) True albedo, (b) Histogram of (a), (c) Normalized unit variance residual image, (d) Histogram of (c).

Under the NMNV assumption,  $C_{\rho}$  is a (non-constant) diagonal matrix. As  $C_{\rho}$  is diagonal,  $C_{\rho\rho^{(0)}}$  is diagonal. Since  $C_{\rho}$  is diagonal,  $C_w$  and thus  $C_{\rho^{(0)}}$  are also diagonal. Therefore, the

LMMSE filtered output (7) simplifies to the following scalar (point) processor of the form

$$\hat{\rho}_{i,j} = E(\rho_{i,j}) + \alpha_{i,j}(\rho_{i,j}^{(0)} - E(\rho_{i,j}^{(0)}))$$

$$\text{where, } \alpha_{i,j} = \frac{\sigma_{i,j}^2(\rho)}{\sigma_{i,j}^2(\rho) + \sigma_{i,j}^2(\mathbf{w})} \quad (12)$$

where  $\sigma_{i,j}^2(\rho)$  and  $\sigma_{i,j}^2(\mathbf{w})$  are the non-stationary signal and noise variances respectively. Recalling that  $E(\rho^{(0)}) = E(\rho)$ , (12) can be written as

$$\hat{\rho}_{i,j} = (1 - \alpha_{i,j})E(\rho_{i,j}) + \alpha_{i,j}\rho_{i,j}^{(0)} \quad (13)$$

So the LMMSE albedo estimate is the weighted sum of the ensemble mean  $E(\rho)$  and the observation  $\rho^{(0)}$ , where the weight depends on the ratio of signal variance to the noise variance. For low signal to noise ratio (SNR) regions, more weight is given to the *a priori* mean  $E(\rho)$  as the observation is too noisy to make an accurate estimate of the original signal. On the other hand, for high SNR regions, more weight is given to the observation.

### C. Noise Variance

From (8), the signal-dependent noise  $\mathbf{w}$  is

$$\mathbf{w}_{i,j} = \frac{(\mathbf{n}_{i,j} - \mathbf{n}_{i,j}^{(0)}) \cdot \mathbf{s} + \mathbf{n}_{i,j}^{(0)} \cdot (\mathbf{s} - \mathbf{s}^{(0)})}{\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}} \rho_{i,j} \quad (14)$$

We assume that the error in surface normal  $(\mathbf{n}_{i,j} - \mathbf{n}_{i,j}^{(0)})$  is uncorrelated in  $x$ ,  $y$  and  $z$  directions and their variances are same. A similar assumption on the error in the light source direction  $(\mathbf{s} - \mathbf{s}^{(0)})$  leads to the following expression for the noise variance  $\sigma^2(\mathbf{w})$

$$\sigma_{i,j}^2(\mathbf{w}) = \sigma_{i,j}^2(\mathbf{w}_1) + \sigma_{i,j}^2(\mathbf{w}_2) \quad (15)$$

where,

$$\sigma_{i,j}^2(\mathbf{w}_1) = \sigma_{i,j}^2(\mathbf{n}) \left( \frac{s_x^2 + s_y^2 + s_z^2}{(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})^2} \right) E(\rho_{i,j}^2) \quad (16)$$

and

$$\sigma_{i,j}^2(\mathbf{w}_2) = \sigma^2(\mathbf{s}) \left( \frac{(n_x^{(0)})^2 + (n_y^{(0)})^2 + (n_z^{(0)})^2}{(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})^2} \right) E(\rho_{i,j}^2) \quad (17)$$

Here  $\sigma_{i,j}^2(\mathbf{n})$  and  $\sigma^2(\mathbf{s})$  are the error variances in each direction of the surface normal and light source direction respectively.  $\{s_x, s_y, s_z\}$  and  $\{n_x^{(0)}, n_y^{(0)}, n_z^{(0)}\}$  are the three components of the illuminant direction and initial surface normal respectively.  $[s_x, s_y, s_z]$  and  $[n_x^{(0)}, n_y^{(0)}, n_z^{(0)}]$  being unit vectors, the expression for the noise variance can further be simplified as follows

$$\sigma_{i,j}^2(\mathbf{w}) = \frac{\sigma_{i,j}^2(\mathbf{n}) + \sigma^2(\mathbf{s})}{(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})^2} E(\rho_{i,j}^2) \quad (18)$$

Appropriately, the noise variance is proportional to the error variances of normal and light source estimates and the variance of the original signal. Interestingly, the noise variance is inversely proportional to  $(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})$  which is the cosine of the angle between the estimates of surface normal and light source direction.

We investigate the correctness of such a relation using a synthetically generated image. Fig. 4 (left) shows the error in

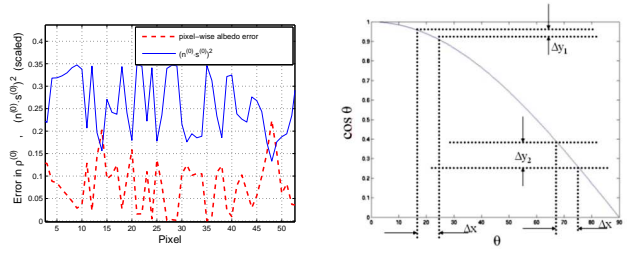


Fig. 4. Left: Pixel-wise albedo error vs  $(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})^2$ . Right: Cosine function explaining the error variation.

$\rho^{(0)}$  for a synthetically generated face image. We see that the error actually varies inversely with  $(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})$  when all the other factors are constant. Such an observation can be attributed to the nature of the cosine function as shown in Fig. 4 (right). When the angle is small, the cosine function changes slowly which implies that small errors in the angle estimate ( $\Delta x$ ) will not adversely affect the accuracy of  $\rho^{(0)}$ , i.e.  $\Delta y_1$  is small. On the other hand, when the angle is large, even a small error in the angle estimate can lead to large errors in  $\rho^{(0)}$ . The noise variance expression used in the proposed estimation framework is capable of accounting for this variation and thus has good potential to obtain a fairly accurate estimate of albedo. The various steps of the proposed algorithm to obtain the albedo estimate from an input intensity image are enumerated in Fig. 5.

### D. Illustration with synthetically generated data

Fig. 6 shows the albedo maps obtained using the proposed algorithm for a face image. To facilitate comparisons with ground truth, the input face image is generated using 3D facial data [2]. Both correct and average facial surface normals are used as  $\mathbf{n}^{(0)}$  to show the efficacy of our approach for a wide range of errors in surface normals. The other contextual information required to obtain the LMMSE estimate of the albedo is determined as follows

- Input: 2D intensity image  $I$  and average surface normal  $\mathbf{n}^{(0)}$ .
- Get initial estimate of source  $\mathbf{s}^{(0)}$  in a least squares manner assuming unit albedo.
- Get initial raw estimate of albedo  $\rho^{(0)}$  using  $\mathbf{n}^{(0)}$  and  $\mathbf{s}^{(0)}$

$$\rho_{i,j}^{(0)} = \frac{I_{i,j}}{\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}}$$

- Estimate the non-stationary noise variance  $\sigma_{i,j}^2(\mathbf{w})$  using

$$\sigma_{i,j}^2(\mathbf{w}) = \frac{\sigma_{i,j}^2(\mathbf{n}) + \sigma^2(\mathbf{s})}{(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})^2} E(\rho_{i,j}^2)$$

- Calculate the LMMSE estimate of the true unknown albedo by linearly combining the signal ensemble average  $E(\rho)$  and the initial raw albedo  $\rho^{(0)}$  as follows

$$\hat{\rho}_{i,j} = (1 - \alpha_{i,j})E(\rho_{i,j}) + \alpha_{i,j}\rho_{i,j}^{(0)}$$

$$\text{where, } \alpha_{i,j} = \frac{\sigma_{i,j}^2(\rho)}{\sigma_{i,j}^2(\rho) + \sigma_{i,j}^2(\mathbf{w})}$$

Fig. 5. Algorithm for finding the LMMSE estimate of albedo

- The illuminant direction  $s^{(0)}$  is estimated using [26].  $\sigma^2(s)$  is estimated by generating a large number of images under randomly selected lighting conditions and estimating their illumination directions.
- $\sigma^2(n)$  is estimated from 3D face data [2]. The data consists of surface normal information for 100 laser-scanned faces.
- Initial albedo  $\rho^{(0)}$  is obtained using (2).
- $E(\rho)$ ,  $\sigma^2(\rho)$  and  $E(\rho^2)$  are estimated from facial albedo data [2]. Fig. 7 shows maps of  $E(\rho)$ ,  $\sigma^2(\rho)$  and  $\sigma^2(n)$ .

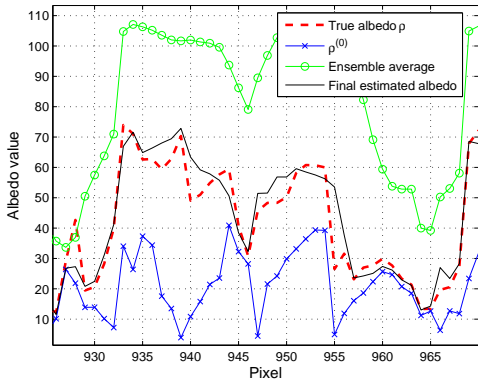


Fig. 6. Estimated albedo maps. Average per-pixel errors are in the ratio (b):(e):(c):(f) :: 17:9:26:12. The plot shows the final albedo estimate as compared to the true albedo, initial albedo  $\rho^{(0)}$  and the ensemble average.

The estimated albedo maps (Fig. 6) seem to be free of *shadowy* effects present in the input image and are quite close to the true albedo map. As zero intensity pixels do not provide any albedo information, a few black regions can be seen in the estimated albedo maps.



Fig. 7. Mean and variance maps used. (a) Ensemble mean of albedo, (b) Ensemble variance of albedo, (c) Error variance of the surface normal.

The improvement in the albedo maps can be explained using the plot in Fig. 6. Though both the ensemble average and the initial albedo  $\rho^{(0)}$  are quite far from the true albedo, their linear combination follows the true value closely. Thus the approach does well in choosing appropriate combining coefficients  $\alpha_{i,j}$  in (13), so that the variation in the accuracy of  $\rho^{(0)}$  at different points is duly accounted for. The improvement obtained over  $\rho^{(0)}$  is significant and is consistent across images of different faces in several different challenging illumination conditions. When tried on 1000 images, the average reduction in per-pixel albedo error is observed to be over 33%. We observe that the improvement in albedo estimates is consistent and is not overly sensitive to the used statistics.

#### IV. SHAPE RECOVERY

In this section, we focus on the general SFS problem of estimating the shape of an object with varying albedo map from a single image. This is an ill-posed problem with too many unknowns and just one constraint per pixel. Traditionally, assumptions like constant/piece-wise constant albedo and known illuminant direction are made to make the problem somewhat tractable. Though important to address the ever-elusive problem of shape recovery from a single image, these assumptions make the SFS approaches ineffective for real objects with varying albedo. In our approach, we transform the original problem of estimating shape of an object with varying albedo map and unknown illumination, to one of estimating the shape of an object with constant albedo and known light source direction that can be addressed using traditional SFS approaches.

We describe in detail each step of the proposed algorithm using a face image as an example. We use 3D information of an average face model as the initial estimate. Using the average shape, we obtain an initial estimate of illuminant direction by formulating it as a linear Least Squares problem [26]

$$s^{(0)} = \left( \sum_{i,j} n_{i,j}^{(0)} n_{i,j}^{(0)T} \right)^{-1} \sum_{i,j} I_{i,j} n_{i,j}^{(0)} \quad (19)$$

where  $n^{(0)}$  is the average facial surface normal. Starting with these initial normal and illuminant estimates, the algorithm proceeds as follows.

##### A. Albedo Estimation and Image Normalization

Given an image, a robust albedo estimate is determined using the image estimation approach described in the preceding section. The albedo estimate  $\hat{\rho}$  is used to normalize the input image to obtain an *albedo-free* image  $G$  as follows

$$G_{i,j} = \frac{I_{i,j}}{\hat{\rho}_{i,j}} \quad (20)$$

The normalized image  $G$  is an image of an object with the same shape as that of the original one but with unit albedo map. Fig. 8 shows an example of the normalized image obtained from a synthetic face image. The normalized image appears quite close to the *true* normalized image obtained directly from the shape information. Also, both the images are quite different from the input image highlighting the importance of such a normalization step before shape estimation.

For a Lambertian object,  $G_{i,j}$  represents the cosine of the angle between the true unknown surface normal  $n_{i,j}$  and the true

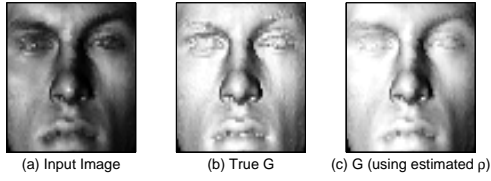


Fig. 8. Normalized image obtained using the albedo estimate.

unknown illuminant direction  $s$  as illustrated in Fig. 10 (right). So using  $s^{(0)}$  to recover the shape may introduce errors in the output depending on the errors in the source estimate. Though the normalized image can potentially be used to estimate the surface normal and refine the illuminant direction estimate using a suitable iterative optimization scheme (e.g., [26]), most traditional SFS approaches assume known light source direction because of possible stability issues in such iterative optimizations. Here, we propose an estimation framework to transform  $G$  to another *albedo-free* image illuminated by a known light source.

### B. Image Transformation

In this step, we transform  $G_{i,j}$  to another image  $H_{i,j}$  that represents the cosine of the angle between the true unknown surface normal  $n_{i,j}$  and the known light source estimate  $s^{(0)}$  (Fig. 10 (right)). An image estimation framework that utilizes the statistics of error in the source estimate is used for this task. The normalized image  $G$  can be written as

$$G = n^T s \quad (21)$$

Now for each pixel, writing the true illuminant direction  $s$  in terms of the initial estimate  $s^{(0)}$  and the difference between the two, we obtain

$$G_{i,j} = n_{i,j} \cdot s^{(0)} + n_{i,j} \cdot (s - s^{(0)}) \quad (22)$$

Identifying that  $H_{i,j}$  represents the cosine of the angle between the true normal  $n_{i,j}$  and the initial estimate of the illuminant direction  $s^{(0)}$ , (22) simplifies to

$$G_{i,j} = H_{i,j} + \nu_{i,j} \quad (23)$$

where  $\nu_{i,j} = n_{i,j} \cdot (s - s^{(0)})$ . As before, this can be viewed as an image estimation problem to obtain an estimate of the transformed image  $H$ . Here the normalized image  $G$  is the degraded signal and  $\nu$  is the observation noise.

Similar to the albedo estimation case, imposing linear constraint on the estimator structure, the LMMSE estimate is given by [27]

$$\hat{H}_{i,j} = (1 - \beta_{i,j})E(H_{i,j}) + \beta_{i,j}G_{i,j} \quad (24)$$

where, 
$$\beta_{i,j} = \frac{\sigma_{i,j}^2(H)}{\sigma_{i,j}^2(H) + \sigma_{i,j}^2(\nu)}$$

Here  $\sigma^2(\nu)$  and  $\sigma^2(H)$  are the non-stationary noise and signal variance respectively. The derivation for the expression for the LMMSE estimate of  $H$  in (24) follows in the same fashion as for albedo and hence omitted for brevity. As before, assuming that the error in the illuminant direction ( $s - s^{(0)}$ ) is uncorrelated in the  $x$ ,  $y$  and  $z$  directions with the same variance  $\sigma^2(s)$ , we have

$$\sigma_{i,j}^2(\nu) = (n_x^2 + n_y^2 + n_z^2)\sigma^2(s) = \sigma^2(s) \quad (25)$$

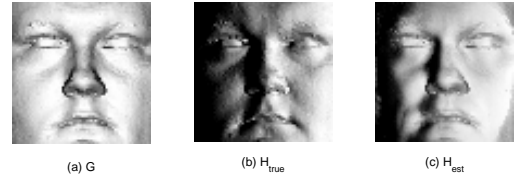


Fig. 9. The transformed image (c) obtained using the proposed estimation framework. For comparison, the normalized image  $G$  and the *true*  $H$  (generated from the 3D data) are also shown. Average per-pixel errors in  $G$  and  $\hat{H}$  are in the ratio of 3:1.

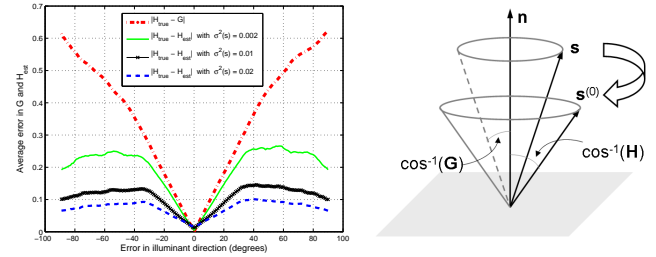


Fig. 10. Left: The variation of error in  $G$  and  $\hat{H}$  with an increase in the angular difference between the estimated and true illuminant direction for different values of  $\sigma^2(s)$ . Right: Image transformation.

Fig. 9 shows the transformed image obtained using this approach. The true source direction in this case is  $[0, 0, -1]$  while  $s^{(0)}$  is taken to be  $[-0.86, 0, -0.52]$ . Clearly, the advantage gained using this image transformation step depends on the error in  $s^{(0)}$ . We perform an experiment to observe the deviation in  $G$  and  $\hat{H}$  from  $H$  with the increase in error in the estimated illuminant direction. As shown in Fig. 10 (left), both  $G$  and  $\hat{H}$  are close to  $H$  when the error in  $s^{(0)}$  is small. The difference  $|H - G|$  increases almost linearly with an increase in the source error. On the other hand, the error in  $\hat{H}$  saturates quickly, highlighting the advantage of the proposed estimation framework to obtain a reliable estimate of  $H$ . The experiment is repeated with different values for the source error variance. The value of  $\sigma^2(s)$  indicates the confidence in the estimated illuminant direction and thus the weight given to the normalized image  $G$  in comparison to the ensemble average  $E(H)$ .

To evaluate the usefulness of the proposed transformation from  $G$  to  $H$ , we estimate the typical errors in light source estimate using 1000 synthetic images and estimating illuminant direction using (19). The average error in the source direction estimate is around  $16^\circ$  indicating approximately 50% reduction (0.145 to 0.07) in average per-pixel error from  $G$  to  $\hat{H}$  for  $\sigma^2(s) = 0.01$ , the value used in our experiments. The different contextual information required to obtain the transformed image is determined from the 3D facial surface normal data [2]. The estimated illuminant direction  $s^{(0)}$  is used for generating a large number of transformed images which are then used for estimating the ensemble mean  $E(H)$  and variance  $\sigma^2(H)$ . Fig. 10 (right) visually demonstrates the image transformation procedure.

Fig. 11 shows the albedo maps and transformed images obtained using the proposed approach on real images from the Yale Face Database B [28]. As desired, the transformed images seem to be less affected due to variations in albedo than the original input images. Note that  $\hat{H}$  is the LMMSE estimate of the image of an object with same shape as the original object but with unit



Fig. 11. The estimated albedo maps  $\hat{\rho}$  and transformed images  $\hat{H}$  obtained for a few subjects from Yale dataset [28].

albedo map and illuminated by the light source  $s^{(0)}$ . Thus one can use a suitable SFS algorithm to solve for the unknown shape.

### C. Shape Estimation

In our implementation, we use the SFS approach by Tsai and Shah [29] that uses a linear approximation of the reflectance function. Here we provide a brief overview of the method for completion. For Lambertian surfaces, the reflectance function  $R$  has the following form

$$R(\mathbf{p}_{i,j}, \mathbf{q}_{i,j}) = \frac{\mathbf{s} \cdot [\mathbf{p}_{i,j}, \mathbf{q}_{i,j}, 1]^T}{\sqrt{1 + \mathbf{p}_{i,j}^2 + \mathbf{q}_{i,j}^2}} \quad (26)$$

where  $\mathbf{p}_{i,j}$  and  $\mathbf{q}_{i,j}$  are the surface gradients. Employing discrete approximations for  $\mathbf{p}$  and  $\mathbf{q}$  using finite differences, we get

$$\begin{aligned} 0 &= f(\hat{H}_{i,j}, \mathbf{Z}_{i,j}, \mathbf{Z}_{i-1,j}, \mathbf{Z}_{i,j-1}) \\ &= \hat{H}_{i,j} - R(\mathbf{Z}_{i,j} - \mathbf{Z}_{i-1,j}, \mathbf{Z}_{i,j} - \mathbf{Z}_{i,j-1}) \end{aligned} \quad (27)$$

where  $\mathbf{Z}_{i,j}$  denotes the depth values. For a given transformed image  $\hat{H}$ , a linear approximation of the function  $f$  about a given depth map  $\mathbf{Z}^{n-1}$  leads to a linear system of equations that can be solved using the Jacobi iterative scheme as follows

$$0 = f(\mathbf{Z}_{i,j}) \approx f(\mathbf{Z}_{i,j}^{n-1}) + (\mathbf{Z}_{i,j} - \mathbf{Z}_{i,j}^{n-1}) \frac{d}{d\mathbf{Z}_{i,j}} f(\mathbf{Z}_{i,j}^{n-1}) \quad (28)$$

Now for  $\mathbf{Z}_{i,j} = \mathbf{Z}_{i,j}^n$ , the depth map at  $n$ -th iteration can be solved using

$$\mathbf{Z}_{i,j}^n = \mathbf{Z}_{i,j}^{n-1} + \frac{-f(\mathbf{Z}_{i,j}^{n-1})}{\frac{d}{d\mathbf{Z}_{i,j}} f(\mathbf{Z}_{i,j}^{n-1})} \quad (29)$$

In our experiments, we use the domain-specific average shape as the initial depth map. Fig. 12 shows the various steps of the proposed albedo estimation and shape recovery algorithm. Depending on the application, one can potentially repeat the sequence of steps with the updated estimates to further refine the albedo, shape and illuminant direction. In our experiments with face images, we do not see much improvement in the estimates after first (or first few) iteration(s). Therefore, all the results shown in the paper are generated using a single parse through the proposed steps. The whole process takes around 2-3 seconds using an unoptimized MATLAB code on a regular desktop.

## V. EXPERIMENTS

We provide details of the experiments performed to evaluate the robustness and usefulness of the albedo and shape estimates obtained using the proposed image estimation framework. We provide examples of albedo estimates obtained on real images. The albedo estimates are also used to relight images under frontal illumination condition. The illumination-insensitivity of the albedo estimates is highlighted by using them to recognize faces across illumination variations. The face recognition application requires albedo estimates from images illuminated by multiple illumination sources. Therefore, the proposed estimation framework is extended to deal with realistic multiple source scenarios.

Other than the image estimation framework to estimate albedo, the main contribution of the work is in transforming the input image to an albedo-free image taken in the presence of known light source which can be used for shape recovery using existing techniques. Therefore, we perform a recognition experiment to evaluate the efficacy of the transformed image before using it to recover 3D shape. The recovered shapes are compared with the available 3D information. The effectiveness of the approach is further highlighted by using the shapes recovered from images downloaded from the web to generate novel views taken under novel illumination.

### A. Illumination-Insensitivity of Estimated Albedo

Fig. 13 shows the albedo maps obtained from several images of a subject from the PIE dataset [30] taken under different illumination conditions. Average facial surface normals are used as  $\mathbf{n}^{(0)}$ . The illuminant direction  $\mathbf{s}^{(0)}$  is estimated using (19). The final albedo estimates obtained using the proposed approach appear much better than the initial erroneous ones and do not seem to have the *shadowy* effects present in the input images. In addition, as desired the estimated albedo maps appear quite similar to each other.

We also perform a relighting experiment using the estimated albedo maps to generate images under frontal illumination. The relighting is performed by combining the estimated albedo maps with average facial shape information. Fig. 14 (second row)

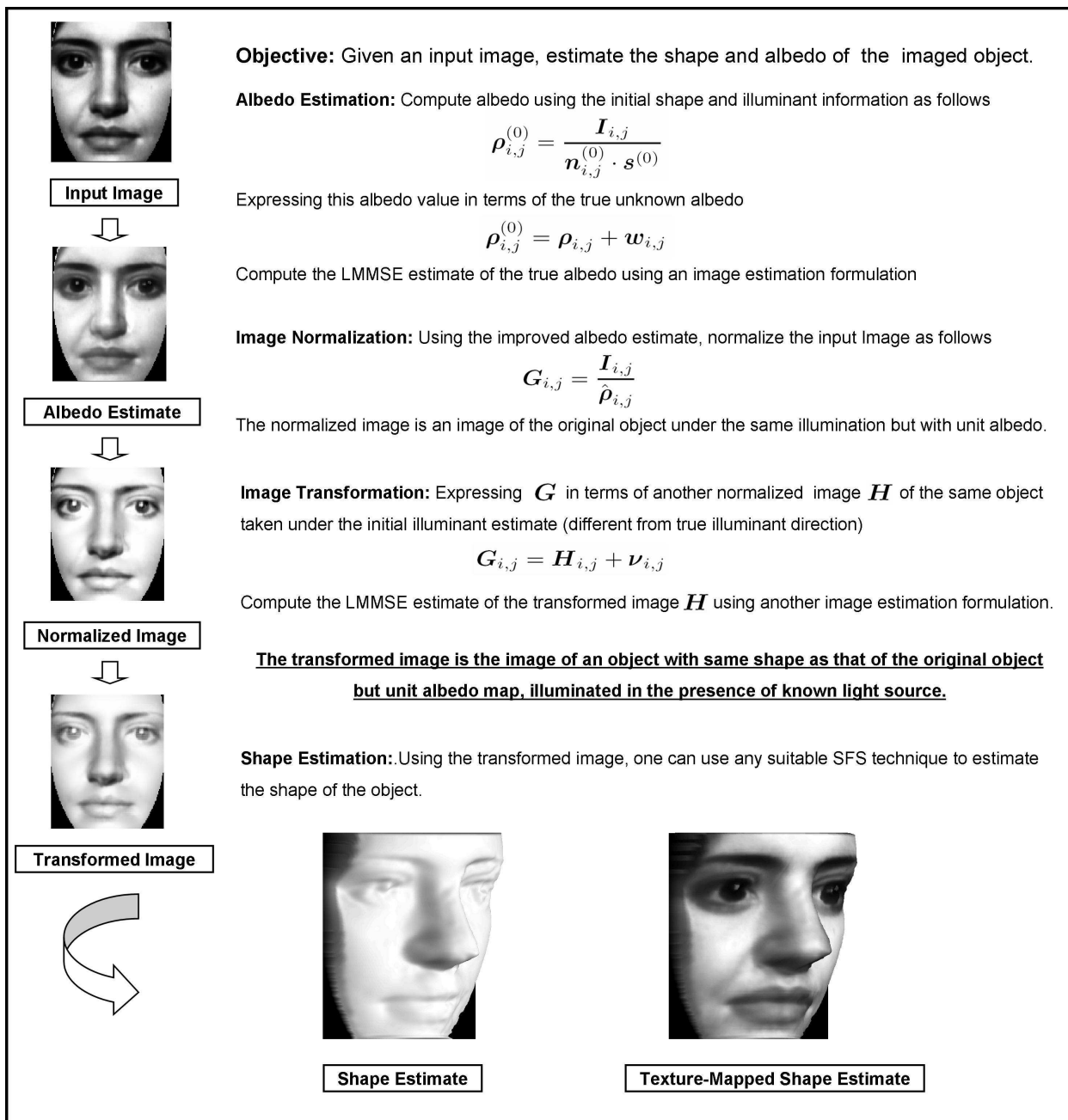


Fig. 12. Visual demonstration of the proposed algorithm.

shows the relighted images and the corresponding input images (first row) taken under challenging illumination conditions. The relighted images seem quite similar to the actual frontally illuminated images of the same subjects from the dataset shown in the third row.

### B. Face Recognition

We now evaluate the usefulness of the estimated albedo maps as an illumination-insensitive signature. We perform recognition experiments on the PIE dataset that contains face images of 68 subjects taken under several different illumination conditions. Given the estimated albedo maps, the similarity between images is measured using Principal Component Analysis (PCA). FRGC [31]

training data consisting of 366 face images is used to generate the (albedo) PCA space. Recognition is performed across illumination with images from one illumination condition forming the gallery while images from another illumination condition forming the probe set. In this experiment setting, each gallery and probe set contains just one image per subject. Table I shows the rank-1 recognition results obtained. Each entry in the table shows the rank-1 performance obtained for one choice of gallery and probe set. The albedo maps perform quite well as illumination-insensitive signatures with an overall average recognition accuracy of 94%.

As shown in the table, the performance compares favorably with those of [14] and [32] which follow similar experimental setting. The performance is also comparable to the one reported



Fig. 13. Albedo estimates obtained from several images of a subject from the PIE dataset [30].



Fig. 14. Relighting results on a few images from the PIE dataset [30].

by Romdhani *et al.* [3] and Zhang and Samaras [1]. Using a 3D morphable model based algorithm, Romdhani *et al.* obtain an average recognition rate of 98% (same as ours with  $f_{12}$  as gallery) using the frontally illuminated images (flash  $f_{12}$ ) as gallery. An average recognition rate of 99% (with  $f_{12}$  as gallery) is reported by Zhang and Samaras using their spherical harmonics based approach. In addition, there are approaches that use/require multiple reference images per identity that have reported recognition results on the PIE dataset. Using multiple reference images including frontally illuminated images, Savvides *et al.* [33] report nearly perfect recognition performance. Yang *et al.* [34] obtain an average recognition score of 94.78% using three reference images. In comparison, we obtain an average recognition score of 98% and 99% with  $f_{12}$  and  $f_{13}$  as gallery respectively. Initial albedo maps  $\rho^{(0)}$  perform poorly in this experiment with an overall average rank-1 performance of 27%. This may be due to the fact that average facial normals are far from the true normals leading to large errors. It is worthwhile to

note that the proposed estimation framework is able to take care of such large errors leading to good recognition performance.

There are quite a few differences between [1], [3], [14], [32] and the proposed approach that warrant some clarifications. First, our experimental setup is restrictive with face images assumed to be in frontal pose (though it is potentially extensible to deal with non-frontal poses). However, there are quite a few advantages that the proposed approach offers as compared to other existing approaches. The proposed algorithm does not involve any costly optimization step and is easily and efficiently implementable. In addition, albedo estimation requires limited domain knowledge in the form of ensemble means and variances. In fact, as discussed in Section V-F, one can replace the ensemble information by local statistics to obtain albedo estimates using the proposed framework. This makes the approach relatively easier to extend to general objects where the domain-dependent statistics is not available. Moreover, the proposed albedo estimation approach does not impose a linear statistical constraint on the unknown

TABLE I

RECOGNITION RESULTS ON THE PIE DATASET [30] USING THE ESTIMATED ALBEDO. WE INCLUDE AVERAGES FROM [14] AND [32] FOR COMPARISON.  $f_i$  DENOTES IMAGES WITH  $i^{th}$  FLASH ON AS LABELED IN PIE. EACH  $(i, j)^{th}$  ENTRY IS THE RANK-1 RECOGNITION RATE OBTAINED WITH THE IMAGES FROM  $f_i$  AS GALLERY AND  $f_j$  AS PROBES.

Probe	$f_{08}$	$f_{09}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$	$f_{16}$	$f_{17}$	$f_{20}$	$f_{21}$	$f_{22}$	Avg	Avg [14]	Avg [32]
Gallery															
$f_{08}$	-	100	100	99	93	91	79	72	44	100	96	85	87	89	92
$f_{09}$	100	-	100	100	99	97	91	90	75	100	99	93	95	93	97
$f_{11}$	100	100	-	100	100	97	88	78	57	100	100	93	92	92	95
$f_{12}$	99	99	100	-	100	100	96	96	87	100	100	97	98	96	98
$f_{13}$	99	99	100	100	-	100	99	99	90	99	100	100	99	98	100
$f_{14}$	97	99	100	100	100	-	99	97	90	100	100	100	98	99	99
$f_{15}$	84	94	88	100	100	100	-	100	99	93	100	100	96	96	97
$f_{16}$	76	97	79	99	100	99	99	-	100	75	99	100	93	91	94
$f_{17}$	53	82	56	90	96	94	94	100	-	54	96	97	83	80	87
$f_{20}$	100	100	100	100	100	100	94	78	57	-	100	99	93	91	95
$f_{21}$	99	99	100	100	100	100	93	94	85	100	-	97	97	96	99
$f_{22}$	90	99	97	100	100	100	100	97	91	97	100	-	97	98	98
Avg	91	97	93	99	99	98	94	91	80	93	99	96	94	-	-
Avg [14]	88	94	93	97	99	99	96	89	75	93	98	98	-	93	-
Avg [32]	90	97	94	99	99	99	98	93	87	95	99	99	-	-	96

albedo and is easily extensible to realistic multiple illumination scenarios (Section V-C).

### C. Albedo Estimation in Multi-source Scenario

Our analysis so far assumes that the image is illuminated by a single light source. However, the assumption does not hold in many realistic scenarios. One of the main challenges in handling multiple light sources is the absence of *a priori* knowledge of the number and placement of the sources. To handle this, we use the result established by Lee *et al.* [35] that an image of an arbitrarily illuminated object can be approximated by a linear combination of the images of the same object in the same pose, illuminated by nine different light sources placed at pre-selected positions. Using this approximation, the image formation equation becomes

$$\mathbf{I} = \sum_{k=1}^9 \gamma_k \mathbf{I}_k \quad \text{where, } \mathbf{I}_k = \rho \max(\mathbf{n} \cdot \mathbf{s}_k, 0) \quad (30)$$

$\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_9\}$  are the pre-specified illumination directions. The following nine illumination directions [35] are used

$$\begin{aligned} \phi &= \{0, 49, -68, 73, 77, -84, -84, 82, -50\}^\circ \\ \theta &= \{0, 17, 0, -18, 37, 47, -47, -56, -84\}^\circ \end{aligned} \quad (31)$$

Since the source directions are pre-specified, the only unknown to estimate the illumination conditions is  $\gamma$ . Given an image  $\mathbf{I}$  of an object and domain-dependent average surface normals  $\mathbf{n}^{(0)}$ ,  $\gamma$  is estimated in a least squares sense as follows

$$\hat{\gamma} = \mathbf{W}^\dagger \mathbf{I} \quad (32)$$

where  $\mathbf{I}$  is the  $N$  dimensional vectorized image and  $\mathbf{W}_{N \times 9}$  is given by

$$\mathbf{W} = \begin{pmatrix} \max(\mathbf{n}_1^{(0)} \cdot \mathbf{s}_1, 0) & \dots & \max(\mathbf{n}_1^{(0)} \cdot \mathbf{s}_9, 0) \\ \max(\mathbf{n}_2^{(0)} \cdot \mathbf{s}_1, 0) & \dots & \max(\mathbf{n}_2^{(0)} \cdot \mathbf{s}_9, 0) \\ \vdots & \ddots & \vdots \\ \max(\mathbf{n}_N^{(0)} \cdot \mathbf{s}_1, 0) & \dots & \max(\mathbf{n}_N^{(0)} \cdot \mathbf{s}_9, 0) \end{pmatrix}$$

The coefficients  $\hat{\gamma}$  can be used to obtain an initial value of the albedo for each pixel as follows

$$\rho_{i,j}^{(0)} = \frac{\mathbf{I}_{i,j}}{\sum_{k=1}^9 \hat{\gamma}_k \max(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}_k, 0)} \quad (33)$$

Robust albedo estimate is obtained by formulating an image estimation problem as follows

$$\rho_{i,j}^{(0)} = \rho_{i,j} + \mathbf{w}_{i,j} \quad (34)$$

where the signal dependent noise  $\mathbf{w}_{i,j}$  is given by

$$\mathbf{w}_{i,j} = \frac{\sum_{k=1}^9 [(\gamma_k - \hat{\gamma}_k) \mathbf{n} \cdot \mathbf{s}_k + \hat{\gamma}_k (\mathbf{n} - \mathbf{n}^{(0)}) \cdot \mathbf{s}_k]}{\sum_{k=1}^9 \hat{\gamma}_k \mathbf{n}^{(0)} \cdot \mathbf{s}_k} \rho$$

Subscripts  $(i, j)'$ s and explicit max operator have been dropped in the above expression for clarity. Similar to the analysis for the single light source case, the NMNV model for the true albedo leads to the following LMMSE albedo estimate

$$\begin{aligned} \hat{\rho}_{i,j} &= (1 - \alpha_{i,j}) E(\rho_{i,j}) + \alpha_{i,j} \rho_{i,j}^{(0)} \\ \text{where, } \alpha_{i,j} &= \frac{\sigma_{i,j}^2(\rho)}{\sigma_{i,j}^2(\rho) + \sigma_{i,j}^2(\mathbf{w})} \end{aligned} \quad (35)$$

Assuming that the errors in estimation of  $\gamma_k$ 's are uncorrelated and have same variance  $\sigma^2(\gamma)$ , the noise variance for each pixel is given by

$$\begin{aligned} \sigma^2(\mathbf{w}) &= \frac{\sum_{k=1}^9 [\sigma^2(\gamma) E[(\mathbf{n} \cdot \mathbf{s}_k)^2] + \sigma^2(\mathbf{n}) \hat{\gamma}_k^2]}{(\sum_{k=1}^9 \hat{\gamma}_k \mathbf{n}^{(0)} \cdot \mathbf{s}_k)^2} E(\rho^2) \\ \text{where, } E[(\mathbf{n} \cdot \mathbf{s}_k)^2] &= (\mathbf{n}^{(0)} \cdot \mathbf{s}_k)^2 + \sigma^2(\gamma) \sigma^2(\mathbf{n}) \end{aligned}$$

Fig. 15 shows the albedo maps obtained using the proposed approach for a face image illuminated by three light sources. Average facial surface normals are used as  $\mathbf{n}^{(0)}$  in this experiment. The proposed multi-source approach provides the best result and has less *shadowy* effects than the others. The single



Fig. 15. Comparison between albedo maps obtained using single and multi-source frameworks. Average per-pixel errors are in the ratio (b):(e):(c):(f)::30:19:22:12. The input image is illuminated by three light sources.

TABLE II

ACCURACY OF ALBEDO ESTIMATES USING SINGLE AND MULTIPLE SOURCE FRAMEWORKS. ENTRIES SHOW AVERAGE PER-PIXEL ALBEDO ERRORS.

	Single source framework	Multiple source framework	Error reduction
$\rho^{(0)}$	35	22	37%
$\hat{\rho}$	23	14	39%
Error reduction	34%	36%	60%

source estimation framework also improves the corresponding initial albedo  $\rho^{(0)}$ , but the result is not as good as the one obtained using the multiple source algorithm. Table II compares the accuracy of various albedo estimates obtained from images illuminated by multiple light sources. As shown, the single light source assumption results in larger errors. Overall improvement in the albedo estimate obtained by the multiple light framework as compared to the initial albedo map obtained under the single light source assumption is about 60%. One thousand images were used to generate the statistics.

#### D. Face Recognition : Multiple Light Sources

In this section, we evaluate the usefulness of the proposed multi-source framework over the single source one when the images are illuminated by several light sources. In the absence of a controlled multi-light source dataset, we generate multi-light source scenarios using the PIE dataset by combining multiple images for each subject. Randomly chosen two, three or four images under different illumination conditions are combined to form twelve different multi-light source scenarios. For the recognition experiment, one image per subject is randomly selected from the twelve illumination conditions for gallery and another one for the probe set. The experiment is repeated 100 times for different random combinations of gallery and probe sets. Fig. 16 shows Cumulative Match Characteristic (CMC) curves comparing the recognition performance obtained using the albedo maps

estimated by the single source framework with that of multiple source framework. Error bars reflect the variations in recognition performance for the different trials. As expected, the multi-source framework significantly outperforms the single-source one. As before, the initial albedo maps  $\rho^{(0)}$  for both single and multiple source frameworks perform poorly in this experiment.

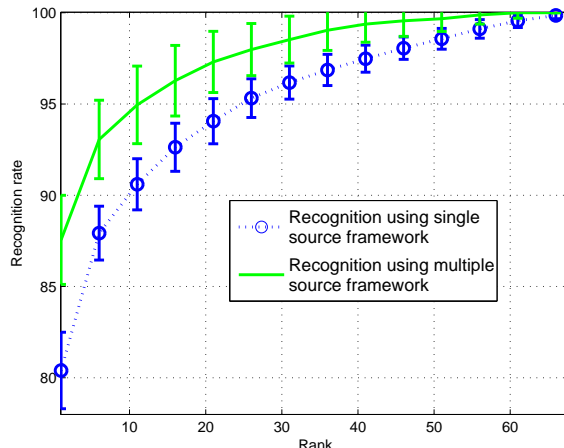


Fig. 16. Recognition performance on *multiply*-illuminated PIE dataset.

#### E. Shape Recovery

As far as the problem of shape recovery is concerned, the main contribution of this work is in generating the *albedo-free* transformed image with known illuminant direction. Therefore, we first evaluate the robustness of the transformed images generated using the proposed approach. Unlike albedo estimates, the transformed images are not illumination-invariant. In fact, a transformed image represents the cosine of the angle between the true unknown surface normal and the known estimate of light source direction, which depends on the illumination in the input image (Fig. 10 (right)). Therefore, unlike albedo, one cannot directly perform a recognition experiment to evaluate the accuracy of the transformed images. Instead, we make use of the statistical facial shape information to derive illumination-insensitive signatures from transformed images. In essence, we force a rank constraint on the unknown shape by writing the transformed image in terms of the basis surface normals as follows

$$H_{i,j} = \max(\mathbf{n}_{i,j} \cdot \mathbf{s}^{(0)}, 0) = \sum_{k=1}^K \mathbf{a}_k \max(\mathbf{n}_{i,j}^k \cdot \mathbf{s}^{(0)}, 0) \quad (36)$$

where  $\mathbf{n}^k$  is the  $k^{th}$  basis surface normal and  $\mathbf{a}_k$  is the corresponding combining coefficient. The coefficient vector  $\mathbf{a}$  being independent of  $\mathbf{s}^{(0)}$ , can be used to perform recognition across illumination variations.

Table III shows the recognition results obtained on the PIE dataset using the coefficient vectors  $\mathbf{a}$  obtained from the corresponding transformed images. The basis normal vectors are derived from the 3D facial data [2] and the coefficient vectors are computed using a closed-form linear least square approach. The overall average recognition rate achieved in this experiment

TABLE III

RECOGNITION RESULTS ON THE PIE DATASET [30] USING THE TRANSFORMED IMAGES.  $f_i$  DENOTES IMAGES WITH  $i^{th}$  FLASH ON AS LABELED IN PIE. EACH  $(i, j)^{th}$  ENTRY IS THE RANK-1 RECOGNITION RATE OBTAINED WITH THE IMAGES FROM  $f_i$  AS GALLERY AND  $f_j$  AS PROBES.

Probe	$f_{08}$	$f_{09}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$	$f_{16}$	$f_{17}$	$f_{20}$	$f_{21}$	$f_{22}$	Avg
Gallery													
$f_{08}$	-	99	99	94	88	74	53	47	26	97	85	57	74
$f_{09}$	94	-	94	99	99	94	71	66	46	93	99	79	85
$f_{11}$	99	99	-	100	99	96	74	57	46	100	100	87	87
$f_{12}$	91	99	100	-	100	100	96	87	71	100	100	99	95
$f_{13}$	87	93	97	100	-	100	99	94	90	96	99	100	96
$f_{14}$	71	96	97	100	100	-	100	99	94	100	100	100	96
$f_{15}$	60	76	75	96	100	100	-	100	100	82	97	100	90
$f_{16}$	41	69	54	90	96	100	100	-	100	62	93	100	82
$f_{17}$	28	44	47	84	93	97	100	100	-	59	88	99	76
$f_{20}$	94	96	100	100	97	96	85	60	57	-	100	91	89
$f_{21}$	85	99	100	100	100	100	97	93	79	100	-	99	96
$f_{22}$	59	84	85	99	100	100	100	100	100	96	99	-	93
Avg	74	87	86	97	97	96	89	82	74	90	96	92	88

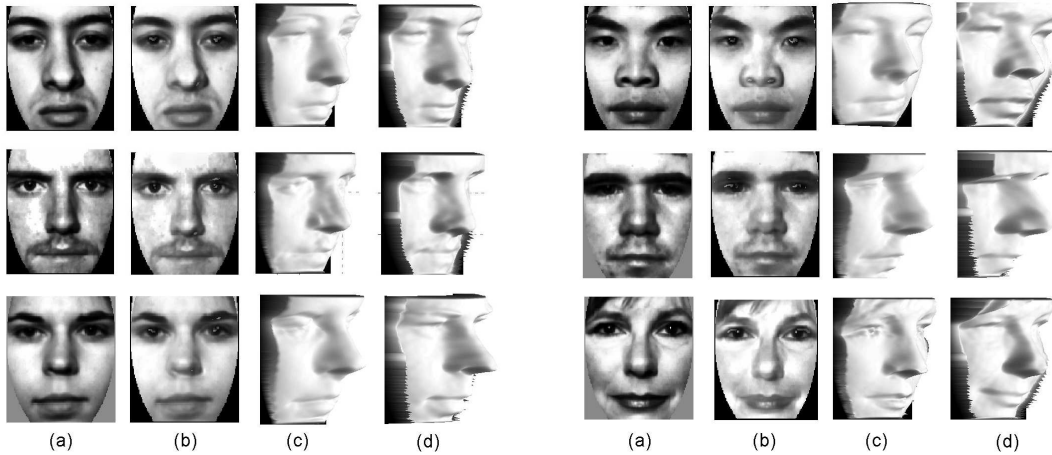


Fig. 17. Comparison with the ground truth. (a) Input image (FRGC dataset), (b) Estimated albedo, (c) Recovered shape, (d) True shape.

is 88%. This signifies the efficacy of the proposed approach in generating robust transformed images.

We now demonstrate the usefulness of the transformed images for the task of shape recovery. Fig. 17 shows a comparison of the recovered shapes with the 3D shapes of the corresponding subjects from the FRGC dataset [31]. The proposed approach seems to recover various person-specific facial features around lips, eyes, etc. Note that the 3D shapes available in the database are captured on a different day than the input intensity images, leading to slightly different facial expressions in the estimated and true shapes.

We perform another experiment to quantitatively evaluate the shape estimates obtained using the transformed images. We compare the shape estimates with the ones obtained using the approach in [29] that directly uses intensity images as input. 1000 synthetically generated images (using Vetter’s 3D face data) are used to determine the angular error in the estimated normal maps for comparison. We observe an improvement of over 15% using our approach.

We also test the efficacy of the approach on images downloaded

from the web with little control over the illumination and other imaging conditions. Fig. 18 shows the albedo and shape estimates obtained along with a few novel views synthesized under novel illumination conditions.

F. Application to general objects

All the experiments in this paper are conducted on faces but the approach is applicable to any domain in general where the required error statistics are available. In the absence of ensemble information, the required means and variances can possibly be approximated by local spatial statistics. Under such an approximation, the LMMSE albedo estimate is given by [25]

$$\hat{\rho}_{i,j} = \bar{\rho}_{i,j} + \alpha_{i,j}(\rho_{i,j}^{(0)} - \bar{\rho}_{i,j}^{(0)})$$

where,  $\alpha_{i,j} = \frac{v_{i,j}^2(\rho)}{v_{i,j}^2(\rho) + \sigma_{i,j}^2(w)}$  (37)

where  $\bar{\rho}_{i,j}$  and  $\bar{\rho}_{i,j}^{(0)}$  are the local spatial means of  $\rho_{i,j}$  and  $\rho_{i,j}^{(0)}$  respectively, and  $v_{i,j}^2(\rho)$  is the local spatial variance of  $\rho_{i,j}$ . The

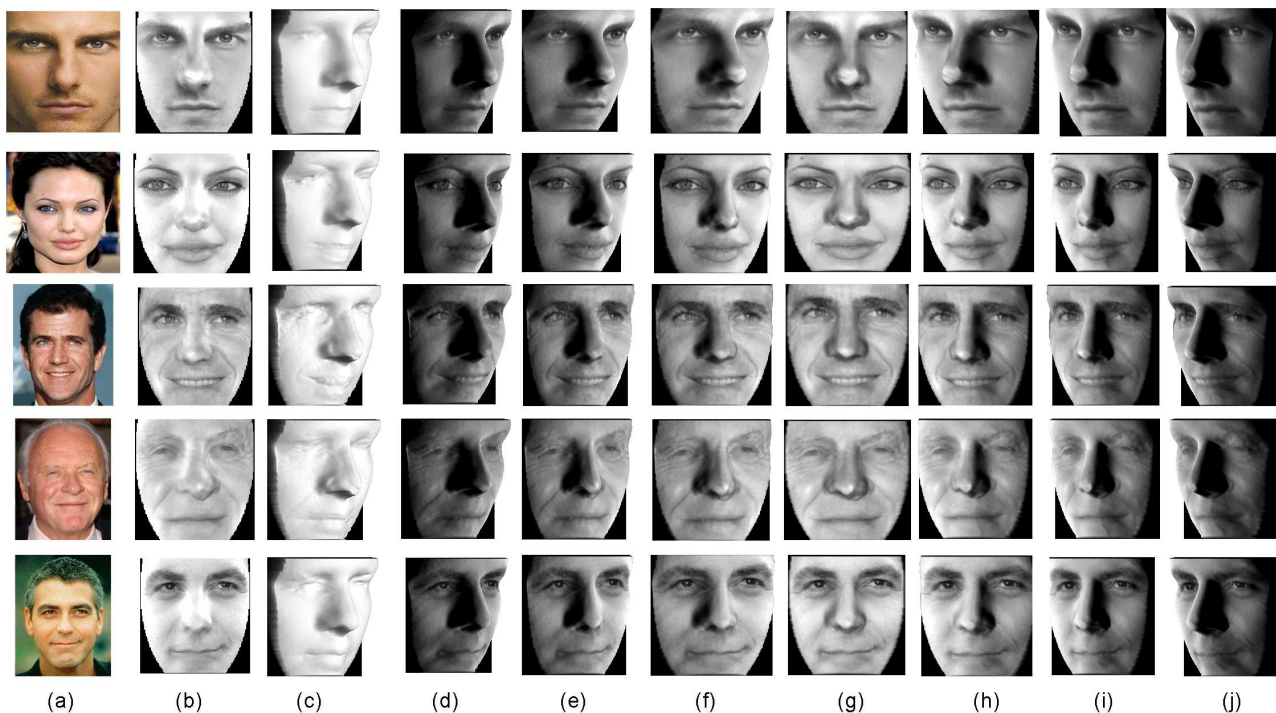


Fig. 18. Novel view synthesis in the presence of novel illumination conditions. (a) Input image, (b) Estimated albedo, (c) Recovered shape, (d)-(j) Synthesized views under novel illumination conditions.



Fig. 19. Top: Images from the Amsterdam Library of Object Images [36]; Middle: Estimated albedo; Bottom: Shadow maps (albedo-free images of the corresponding input images).

local statistics of  $\rho_{i,j}$  can be calculated from that of the degraded signal using the following relations

$$\bar{\rho}_{i,j} = \bar{\rho}_{i,j}^{(0)} \quad (38)$$

$$v_{i,j}^2(\rho) = \frac{v_{i,j}^2(\rho^{(0)}) - (\bar{\rho}_{i,j}^{(0)})^2 A_{i,j}}{1 + A_{i,j}}, \quad (39)$$

where  $A_{i,j} = \frac{\sigma_{i,j}^2(\mathbf{n}) + \sigma^2(\mathbf{s})}{(\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)})^2}$

Though not too accurate, these approximations are probably the best one can do in the absence of any ensemble information. Fig. 19 (middle row) shows the albedo estimates obtained this

way for a few images (top row) from the Amsterdam Library of Object Images [36]. As desired, the illumination effects present in the input images are less visible in the albedo maps. This is further highlighted in the shadow maps shown in Fig. 19 (bottom row). In these examples, the surface normals required to obtain the initial albedo map are assigned manually. For example, the cylindrical shape is used for the mug and the coke can while the cuboidal shape is used for the boxes shown in the figure. Fig. 20 shows some more examples of the albedo estimates obtained along with a few zoomed-in regions to signify the usefulness of the approach for general objects.

## VI. SUMMARY

We proposed an image estimation formulation for the task of albedo estimation from a single image. Errors in illumination and surface normal information lead to erroneous albedo maps. The proposed estimation framework effectively utilizes the statistics of error in illumination and normal information for robust estimation of albedo. The framework is extended to multiple illuminant scenarios to be applicable to images illuminated by an arbitrary number of unknown light sources. Comparisons with true albedo maps are shown to highlight the effectiveness of the algorithm. Extensive experiments are performed to show the usefulness of the estimated albedo maps as illumination-insensitive signatures. The albedo maps are also used to obtain *albedo-free* images for shape recovery. Another estimation framework is developed to reduce the general problem of recovering shape of an object with varying unknown albedo and unknown illuminant direction to the one that can be addressed by the traditional SFS approaches. The recovered shapes and albedo maps are used to synthesize novel views under novel illumination conditions.

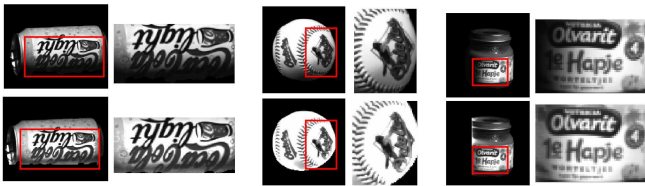


Fig. 20. Top row: Original images of a few objects; Bottom row: Estimated albedo maps obtained using local statistics. The zoomed in regions are shown to highlight the difference between the input images and the corresponding albedo estimates.

## REFERENCES

- [1] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 351–363, March 2006.
- [2] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, Sep 2003.
- [3] S. Romdhani, V. Blanz, and T. Vetter, "Face identification by fitting a 3d morphable model using linear shape and texture error functions," in *Proceedings of European Conference on Computer Vision*, 2002, pp. 3–19.
- [4] W. A. P. Smith and E. R. Hancock, "Recovering facial shape using a statistical model of surface normal direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1914–1930, Dec 2006.
- [5] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 218–233, Feb 2003.
- [6] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of convex Lambertian object," *Journal of the Optical Society of America A*, pp. 2448–2459, Oct 2001.
- [7] B. K. P. Horn and M. J. Brooks, *Shape from Shading*. Cambridge Massachusetts: MIT Press, 1989.
- [8] R. Zhang, P. Tsai, J. Cryer, and M. Shah, "Shape from shading: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690–706, Aug 1999.
- [9] S. Biswas, G. Aggarwal, and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," in *Proceedings of IEEE Conference on Computer Vision*, 2007.
- [10] W. Zhao and R. Chellappa, "Symmetric shape from shading using self-ratio image," *International Journal of Computer Vision*, vol. 45, no. 1, pp. 55–75, Oct 2001.
- [11] J. J. Atick, P. A. Griffin, and A. N. Redlich, "Statistical approach to sfs: Reconstruction of 3d face surfaces from single 2d images," *Neural Computation*, vol. 8, pp. 1321–1340, Aug 1996.
- [12] R. Dovgand and R. Basri, "Statistical symmetric shape from shading for 3d structure recovery of faces," in *Proceedings of European Conference on Computer Vision*, 2004, pp. 99–113.
- [13] D. Samaras and D. Metaxas, "Illumination constraints in deformable models for shape and light direction estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 247–264, Feb 2003.
- [14] S. Zhou, R. Chellappa, and D. Jacobs, "Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints," in *Proceedings of European Conference on Computer Vision*, 2004, pp. 588–601.
- [15] Z. Yue, W. Zhao, and R. Chellappa, "Pose-encoded spherical harmonics for face recognition and synthesis using a single image," *To appear in EURASIP Journal on Advances in Signal Processing*, 2008.
- [16] K. C. Lee and B. Moghaddam, "A practical face relighting method for directional lighting normalization," in *Proceedings of International Workshop on Analysis and Modeling of Faces and Gestures*, Oct 2005.
- [17] A. Hurlbert, "Formal connections between lightness algorithms," *Journal of the Optical Society of America A*, vol. 3, no. 10, pp. 1684–1693, October 1986.
- [18] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Prentice-Hall signal processing series, 1977.
- [19] B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for gaussian models of images," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-6, pp. 876–881, Dec 1976.
- [20] H. T. Trussell and B. R. Hunt, "Sectioned methods for image restoration," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. ASSP-26, pp. 157–164, Apr 1978.
- [21] D. S. Lebedev and L. I. Mirkin, "Smoothing of two-dimensional images using the 'composite' model of a fragment," *Iconics-Digital Holography-Image Processing*, pp. 57–62, 1975.
- [22] G. L. Anderson and A. N. Netravali, "Image restoration based on subjective criterion," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-6, pp. 845–853, Dec 1976.
- [23] J. F. Abramatic and L. M. Silverman, "Nonlinear restoration of noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, pp. 141–149, March 1982.
- [24] F. Naderi and A. A. Sawchuk, "Estimation of images degraded by film-grain noise," *Applied Optics*, vol. 17, pp. 1228–1237, Apr 1978.
- [25] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 165–177, March 1985.
- [26] M. J. Brooks and B. K. P. Horn, "Shape and source from shading," in *Proceedings of International Joint Conference on Artificial Intelligence*, Aug 1985, pp. 932–936.
- [27] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control*. McGraw-Hill, 1971.
- [28] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, June 2001.
- [29] P. S. Tsai and M. Shah, "Shape from shading using linear approximation," *Image and Vision Computing Journal*, vol. 12, no. 8, pp. 487–498, Oct 1994.
- [30] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, Dec 2003.
- [31] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 947–954.
- [32] G. Aggarwal and R. Chellappa, "Face recognition in the presence of multiple light sources," in *Proceedings of International Conference on Computer Vision*, 2005, pp. 1169–1176.
- [33] M. Savvides, B. V. K. V. Kumar, and P. K. Khosla, "Corefaces - robust shift invariant pca based correlation filter for illumination tolerant face recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2004, pp. II: 834–841.
- [34] C. T. Yang, S. Lai, and L. Chang, "Robust face image matching under illumination variations," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 16, pp. 2533–2543, 2004.
- [35] K. C. Lee, J. Ho, and D. Kriegman, "Nine points of light: acquiring subspaces for face recognition under variable lighting," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Dec 2001, pp. 519–526.
- [36] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, Jan 2005.