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

Soma Biswas†, Gaurav Aggarwal†† and Rama Chellappa†
Center for Automation Research, UMIACS
†Dept. of ECE, ††Dept. of Computer Science
University of Maryland, College Park
{soma, gaurav, rama}@cfar.umd.edu

Abstract

In this paper, we propose a non-stationary stochastic filtering framework for the task of albedo estimation from a single image. There are several approaches in literature for albedo estimation, but few include the errors in estimates of surface normals and light source directions to improve the albedo estimate. The proposed approach effectively utilizes the error statistics of surface normals and illumination direction for robust estimation of albedo. The albedo estimate obtained is further used to generate albedo-free normalized images for recovering the shape of an object. Illustrations and experiments are provided to show the efficacy of the approach and its application to illumination-invariant matching and shape recovery.

1. Introduction

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. Albedo estimation has numerous applications in the field of Computer Vision and Graphics. Unlike image intensity, albedo is invariant to changes in illumination conditions which makes it useful for illumination-invariant matching of objects. In graphics, accurate albedo estimates are important for realistic Image based Rendering (IBR).

Albedo estimation has often been coupled with shape estimation. There are methods that simultaneously estimate the shape and albedo of an object [4][12]. A few others first obtain a shape estimate followed by albedo estimation [17]. Then there are methods whose main goal is shape estimation and albedo is finally incorporated to account for the image reconstruction errors using the estimated shape [15]. 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 error in 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 is written 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 robustness of the estimated albedo maps allow us to use them for the task of shape recovery. Traditional Shape-from-Shading (SFS) approaches [7] often make constant/piece-wise constant albedo assumption to recover shape of an object from a single image. Such assumptions though useful to make the problem less intractable, often limit the applicability of the approaches for real world objects with varying albedo. Here, using the estimated albedo, we obtain *albedo-free* images that can be handled by traditional SFS approaches.

Experiments show that albedo estimates obtained using the proposed approach are quite close to the true value. We evaluate the illumination-insensitivity of the estimated albedo maps by using them for face recognition. Recognition results on PIE dataset [14] show that not only the estimated albedo maps significantly outperform the initial erroneous ones but also compares well with several recent illumination-invariant approaches for face recognition.

*Partially funded by UNISYS contract

978-1-4244-1631-8/07/$25.00 ©2007 IEEE

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.

We also show shape recovery results using the proposed approach. The recovered shapes are compared with the ground truth and are also used to synthesize novel views under novel illumination scenarios.

The rest of the paper is organized as follows. Section 2 discusses a few related works. The proposed albedo estimation framework is detailed in Section 3. Section 4 details the steps involved in shape recovery using the estimated albedo. Experimental results are presented in Section 5. Section 6 concludes the paper with a brief summary.

2. Previous Work

Albedo estimation is often seen as a simultaneous problem with shape estimation. An example in this category is the SFS [7] approach. Dealing with an inherently ill-posed problem, SFS research typically makes simplifying assumptions like constant or piece-wise constant albedo. Advances have been made that use domain specific constraints to allow for analysis and estimation of general albedo maps. Zhao and Chellappa [18] recover shape and albedo for the class of bilaterally symmetric objects with a piece-wise constant albedo field. Dovgard and Basri [6] combine the symmetrical SFS formulation with statistical approach to recover shape from a single image. Smith and Hancock [15] 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.

Blanz and Vetter [4] propose a 3D morphable model in which a face is represented using a linear combination of basis exemplars. The shape and albedo parameters of the model are obtained by fitting the morphable model to the input image. Romdhani *et al.* [12] provide an efficient and robust algorithm for fitting a 3D morphable model using shape and texture error functions. Zhang and Samaras [17] combine spherical harmonics illumination representation [3] with 3D morphable models [4]. After obtaining a shape estimate, albedo and illumination coefficients are computed in an iterative fashion. Zhou *et al.* [19] impose a rank constraint on shape and albedo in a class to separate the two from illumination using factorization approach. Lee and Moghaddam [10] propose a scheme for albedo estimation and relighting of human faces, using an average face to determine the dominant light source direction followed by albedo estimation. The problem of albedo estimation has also been addressed by lightness algorithms that recover an approximation to surface reflectance in independent wavelength channels [8].

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 [2], we provide pointers only to a few papers which are useful to understand the statistical machinery used in our non-stationary Wiener filter based formulation. The standard Wiener filter is known to be optimal for second order stationary processes while dealing with additive noise. There are several modifications to the standard Wiener filter to account for non-stationary processes. Anderson and Netravali [1] propose a masking function based on the observation that the eyes tend to be more robust to noise at the edges of the image than at the flat regions. A different Wiener filter is computed for each pixel depending upon this apparent noise variance at the pixel. Kuan *et al.* [9] design a scalar filter for estimating images with signal-dependent noise using the non-stationary mean non-stationary variance model.

3. Albedo Estimation

The surface normals, albedo and the intensity image are related by an image formation model. For Lambertian objects, the diffused component of the surface reflection is modeled using the *Lambert's Cosine Law* given by

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

where I is the pixel intensity, $\boldsymbol{s}$ is the light source direction, ρ is the surface albedo and $\boldsymbol{n}$ is the surface normal of the corresponding 3D point. The expression implicitly assumes a single dominant light source placed at infinity. (The theoretical analysis we propose here is extensible to multiple light source scenarios as discussed in the supplemental material.) 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.

The Lambertian assumption imposes the following constraint on the initial albedo $\boldsymbol{\rho}^{(0)}$ obtained using the initial surface normal $\boldsymbol{n}_{i,j}^{(0)}$ and illuminant direction $\boldsymbol{s}^{(0)}$

$$\boldsymbol{\rho}_{i,j}^{(0)} = \frac{\boldsymbol{I}_{i,j}}{\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}} \tag{2}$$

where $\cdot$ is the standard dot product operator. Clearly, more accurate the denominator $(\boldsymbol{n}^{(0)} \cdot \boldsymbol{s}^{(0)})$ is, the closer is the obtained initial albedo to its true value $\boldsymbol{\rho}$. For most applications, accurate initial estimates of surface normal and light source direction are not available leading to erroneous $\boldsymbol{\rho}^{(0)}$.

Figure 1 illustrates the errors in the obtained albedo $\boldsymbol{\rho}^{(0)}$ for a synthetically generated face image. True surface normal information is used in this example while light source direction is estimated using the method in [5]. One may expect the errors to be larger if inaccurate estimates of surface normals are used. Interestingly, not only is $\boldsymbol{\rho}^{(0)}$ quite far from the true value for quite a few points, but even the error varies appreciably across pixels. The proposed estimation framework duly accounts for these variations to obtain a robust albedo estimate.

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.

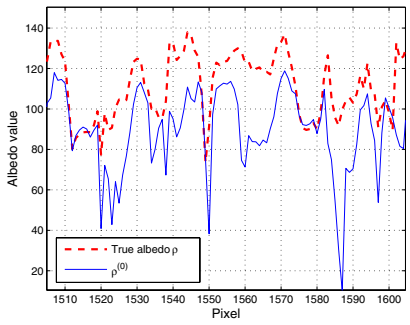

Figure 1. Errors in the initial albedo map $\boldsymbol{\rho}^{(0)}$.

3.1. 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

$$\boldsymbol{\rho}_{i,j}^{(0)} = \frac{\boldsymbol{I}_{i,j}}{\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}} = \boldsymbol{\rho}_{i,j} \frac{\boldsymbol{n}_{i,j} \cdot \boldsymbol{s}}{\boldsymbol{n}_{i,j}^{(0)} \boldsymbol{s}^{(0)}} \tag{3}$$

where $\boldsymbol{\rho}$, $\boldsymbol{n}$ and $\boldsymbol{s}$ are the true unknown albedo, normal and illuminant direction respectively. We suppress the explicit $\max$ operator by considering only the pixels with positive intensity values. $\boldsymbol{\rho}^{(0)}$ can further be expressed as follows

$$\boldsymbol{\rho}_{i,j}^{(0)} = \boldsymbol{\rho}_{i,j} + \frac{\boldsymbol{n}_{i,j} \cdot \boldsymbol{s} - \boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}}{\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}} \boldsymbol{\rho}_{i,j} \tag{4}$$

Substituting $\boldsymbol{w}_{i,j} = \frac{\boldsymbol{n}_{i,j} \cdot \boldsymbol{s} - \boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}}{\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}} \boldsymbol{\rho}_{i,j}$, (4) simplifies to

$$\boldsymbol{\rho}_{i,j}^{(0)} = \boldsymbol{\rho}_{i,j} + \boldsymbol{w}_{i,j} \tag{5}$$

This can be identified with the standard image estimation formulation [2]. Here $\boldsymbol{\rho}$ is the original signal (true albedo), the initial estimate $\boldsymbol{\rho}^{(0)}$ is the degraded signal and $\boldsymbol{w}$ is the signal dependent additive noise.

3.2. LMMSE Estimate of Albedo

The Minimum Mean Square Error (MMSE) estimate of the albedo map $\boldsymbol{\rho}$ given noisy observed map $\boldsymbol{\rho}^{(0)}$ is the conditional mean $\hat{\boldsymbol{\rho}} = E(\boldsymbol{\rho}|\boldsymbol{\rho}^{(0)})$, which depends on the probability density functions of $\boldsymbol{\rho}$ and $\boldsymbol{w}$ and is difficult to estimate. Imposing linear constraint on the estimator structure, the LMMSE estimate is given by [13]

$$\hat{\boldsymbol{\rho}} = E(\boldsymbol{\rho}) + \mathrm{C}_{\rho\rho^{(0)}} \mathrm{C}_{\rho^{(0)}}^{-1} (\boldsymbol{\rho}^{(0)} - E(\boldsymbol{\rho}^{(0)})) \tag{6}$$

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

In conventional image estimation problems, the original signal is assumed to be wide-sense stationary which may be an oversimplified assumption in many situations. Here we assume a Non-stationary Mean Non-stationary Variance (NMNV) model for the original signal $\boldsymbol{\rho}$ which has been shown to be a reasonable assumption for many applications [9]. Under this model, the original signal is characterized by a non-stationary mean and a diagonal covariance matrix C_{ρ} with non-stationary variance.

Signal-dependent noise $\boldsymbol{w}$ can be rewritten as follows

$$\boldsymbol{w}_{i,j} = \frac{(\boldsymbol{n}_{i,j} - \boldsymbol{n}_{i,j}^{(0)}) \cdot \boldsymbol{s} + \boldsymbol{n}_{i,j}^{(0)} \cdot (\boldsymbol{s} - \boldsymbol{s}^{(0)})}{\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}} \boldsymbol{\rho}_{i,j} \tag{7}$$

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

$$\begin{aligned} \mathrm{C}_{\rho\rho^{(0)}} &= E[(\boldsymbol{\rho} - E(\boldsymbol{\rho}))(\boldsymbol{\rho}^{(0)} - E(\boldsymbol{\rho}^{(0)}))^T] \\ &= \mathrm{C}_{\rho} + E[(\boldsymbol{\rho} - E(\boldsymbol{\rho}))\boldsymbol{w}^T] \end{aligned} \tag{8}$$

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

$$\begin{aligned} \mathrm{C}_{\rho^{(0)}} &= E[(\boldsymbol{\rho}^{(0)} - E(\boldsymbol{\rho}^{(0)}))(\boldsymbol{\rho}^{(0)} - E(\boldsymbol{\rho}^{(0)}))^T] \\ &= \mathrm{C}_{\rho} + \mathrm{C}_{w} + E[(\boldsymbol{\rho} - E(\boldsymbol{\rho}))\boldsymbol{w}^T] \\ &+ E[\boldsymbol{w}(\boldsymbol{\rho} - E(\boldsymbol{\rho}))^T] \end{aligned} \tag{9}$$

Recalling that $E(\boldsymbol{w})$ and $E(\boldsymbol{w}|\boldsymbol{\rho})$ are zero, $E\big((\boldsymbol{\rho} - E(\boldsymbol{\rho}))\boldsymbol{w}^T\big) = 0$. This simplifies (8) and (9) as follows

$$\mathrm{C}_{\rho\rho^{(0)}} = \mathrm{C}_{\rho} \quad \text{and} \quad \mathrm{C}_{\rho^{(0)}} = \mathrm{C}_{\rho} + \mathrm{C}_{w} \tag{10}$$

As C_{ρ} is diagonal, $\mathrm{C}_{\rho\rho^{(0)}}$ is diagonal. Furthermore, assuming the noise $\boldsymbol{w}$ is uncorrelated, C_{w} and thus $\mathrm{C}_{\rho^{(0)}}$ are also diagonal. Therefore, the LMMSE filtered output (6) simplifies to the following scalar (point) processor of the form

$$\begin{aligned} \hat{\boldsymbol{\rho}}_{i,j} &= E(\boldsymbol{\rho}_{i,j}) + \boldsymbol{\alpha}_{i,j}\big(\boldsymbol{\rho}_{i,j}^{(0)} - E(\boldsymbol{\rho}_{i,j}^{(0)})\big) \\ \text{where,} \quad \boldsymbol{\alpha}_{i,j} &= \frac{\sigma_{i,j}^2(\boldsymbol{\rho})}{\sigma_{i,j}^2(\boldsymbol{\rho}) + \sigma_{i,j}^2(\boldsymbol{w})} \end{aligned} \tag{11}$$

where $\sigma_{i,j}^2(\boldsymbol{\rho})$ and $\sigma_{i,j}^2(\boldsymbol{w})$ are the non-stationary signal and noise variance respectively. Recalling that $E(\boldsymbol{\rho}^{(0)}) = E(\boldsymbol{\rho})$, (11) can equivalently be written as

$$\hat{\boldsymbol{\rho}}_{i,j} = (1 - \boldsymbol{\alpha}_{i,j})E\big(\boldsymbol{\rho}_{i,j}\big) + \boldsymbol{\alpha}_{i,j}\boldsymbol{\rho}_{i,j}^{(0)} \tag{12}$$

So the LMMSE albedo estimate is the weighted sum of the ensemble mean $E(\boldsymbol{\rho})$ and the observation $\boldsymbol{\rho}^{(0)}$, where the

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.

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 $E(\boldsymbol{\rho})$ 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.

3.3. Noise Variance

From (7), the signal-dependent noise $\boldsymbol{w}$ is

$$\boldsymbol{w}_{i,j} = \frac{(\boldsymbol{n}_{i,j} - \boldsymbol{n}_{i,j}^{(0)}) \cdot \boldsymbol{s} + \boldsymbol{n}_{i,j}^{(0)} \cdot (\boldsymbol{s} - \boldsymbol{s}^{(0)})}{\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}} \boldsymbol{\rho}_{i,j} \qquad (13)$$

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

$$\sigma_{i,j}^2(\boldsymbol{w}) = \frac{\sigma_{i,j}^2(\boldsymbol{n}) + \sigma^2(\boldsymbol{s})}{\left(\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}\right)^2} E\left(\boldsymbol{\rho}_{i,j}^2\right) \qquad (14)$$

where $\sigma_{i,j}^2(\boldsymbol{n})$ and $\sigma^2(\boldsymbol{s})$ are the error variances in each direction of the normal and source direction respectively.

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 $(\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)})$ 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. Figure 2 (left) shows that the error in $\boldsymbol{\rho}^{(0)}$ actually varies inversely with $(\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)})$. Such an observation can be attributed to the nature of the cosine function as shown in Figure 2 (right). When the angle is small, the cosine function changes slowly which implies that small errors in the angle estimate will not adversely affect the accuracy of $\boldsymbol{\rho}^{(0)}$. On the other hand, when the angle is large, even a small error in the angle estimate can lead to large errors in $\boldsymbol{\rho}^{(0)}$. The noise variance expression used in the estimation framework is capable of accounting for this variation and thus has good potential to obtain a fairly accurate estimate of albedo.

Figure 3 shows the albedo maps obtained using the proposed algorithm for a face image. To facilitate comparison with ground truth, the input face image is generated using 3D facial data [4]. Both correct and average facial surface normals are used as $\boldsymbol{n}^{(0)}$ 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

- The illuminant direction $s^{(0)}$ is estimated using [5]. $\sigma^2(\boldsymbol{s})$ is estimated by generating a large number of images under randomly selected lighting conditions and estimating their illumination directions.
- $\sigma^2(\boldsymbol{n})$ is estimated from 3D face data [4].
- Initial albedo $\boldsymbol{\rho}^{(0)}$ is obtained using (2).
- $E(\boldsymbol{\rho})$, $\sigma^2(\boldsymbol{\rho})$ and $E(\boldsymbol{\rho}^2)$ are estimated from facial albedo data [4].

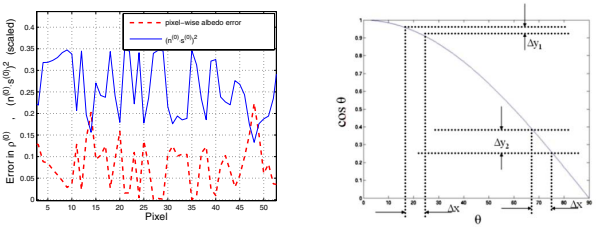

Figure 2. Left: Pixel-wise albedo error vs $\left(\boldsymbol{n}_{i,j}^{(0)} \cdot \boldsymbol{s}^{(0)}\right)^2$. Right: Cosine function explaining the error variation.

The estimated albedo maps (Figure 3) 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.

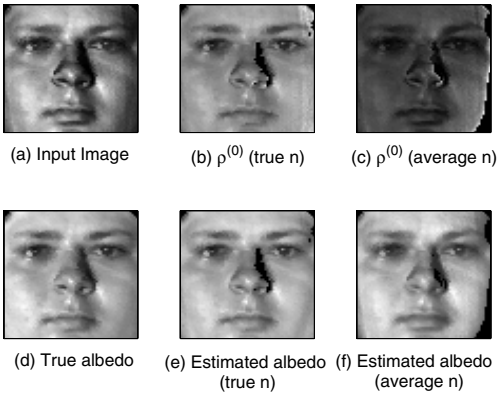

Figure 3. Estimated albedo maps. Average per-pixel errors are in the ratio (b):(e):(c):(f) :: 17:9:26:12.

The improvement in the albedo maps can be explained using the plot in Figure 4. As shown, the approach does well in choosing appropriate combining coefficients $\boldsymbol{\alpha}_{i,j}$ in (12), so that the variation in the accuracy of $\boldsymbol{\rho}^{(0)}$ at different points is duly accounted for. The improvement obtained over $\boldsymbol{\rho}^{(0)}$ 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%.

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.

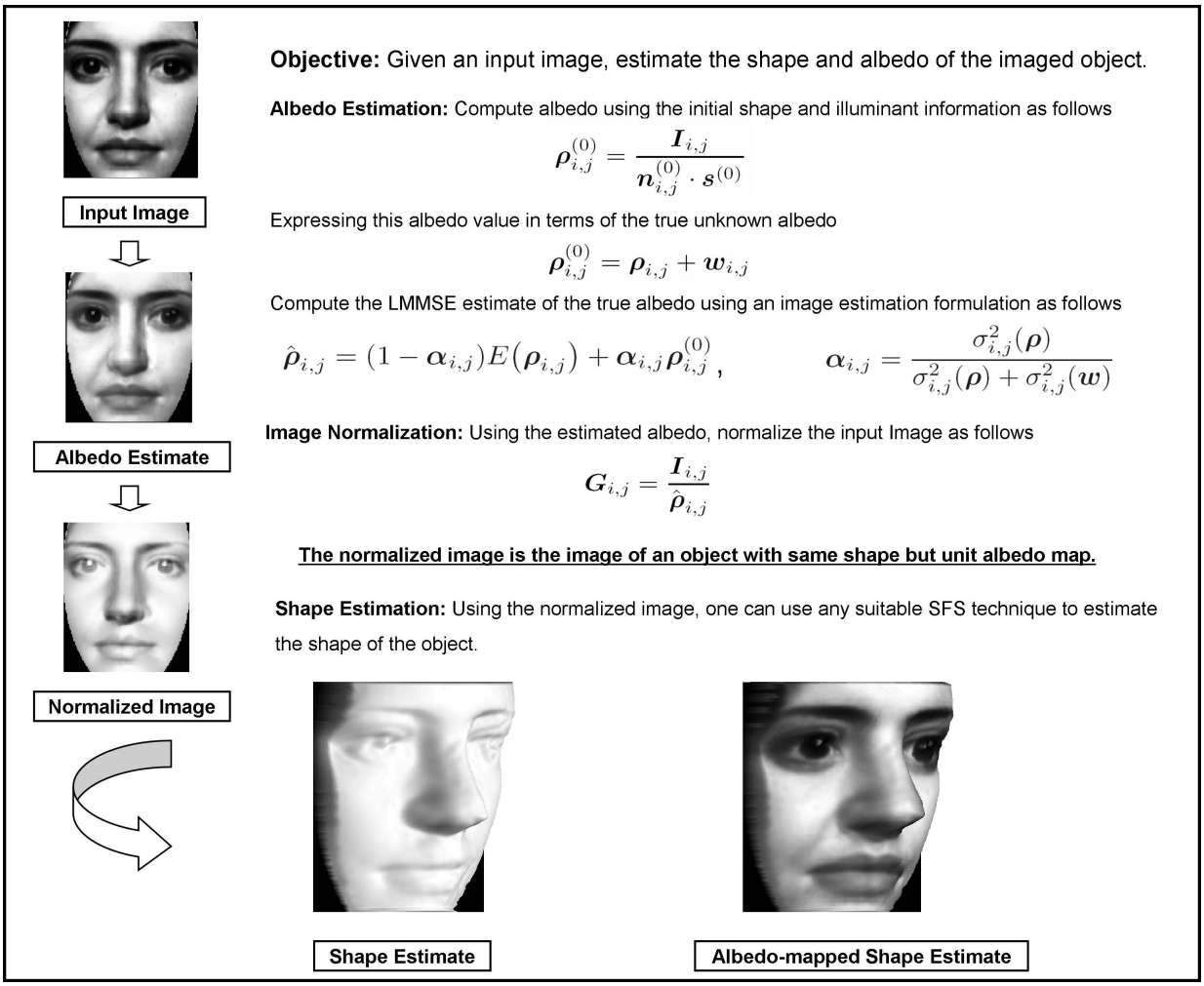

Figure 5. Visual demonstration of the proposed albedo estimation and shape recovery algorithm.

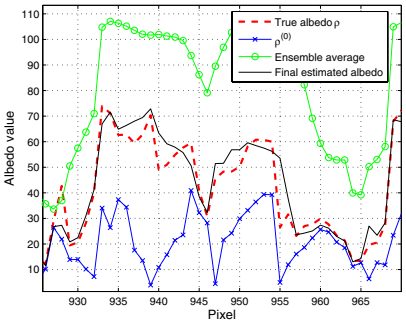

Figure 4. Final albedo estimate as compared to the true albedo, initial albedo $\rho^{(0)}$ and the ensemble average.

4. Shape Recovery

In this section, we focus on the general SFS problem of estimating 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 are made to simplify the problem. Here, we use the estimated albedo maps to transform the original problem to one of estimating shape of an object with constant albedo that can be addressed using a traditional SFS approach.

The albedo estimate $\hat{\rho}$ is used to normalize the input image to obtain an *albedo-free* image G as follows

$$G_{i,j} = \frac{I_{i,j}}{\hat{\rho}_{i,j}} \tag{15}$$

The normalized image G is an image of an object with same shape as that of the original one but with unit albedo map. Figure 6 shows an example of normalized image obtained from a synthetic face image [4]. The normalized image appears quite close to the *true* normalized image obtained directly from the shape information. Also, both the normalized images are quite different from the input image

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.

Figure 6. Normalized image obtained using the albedo estimate.

highlighting the importance of such a normalization step before shape estimation.

Given the *albedo-free* image $\boldsymbol{G}$, one can use a suitable SFS approach for shape recovery. In our implementation, we use the one proposed by Tsai and Shah [16] 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 $\boldsymbol{R}$ has the following form

$$\boldsymbol{R}(\boldsymbol{p}_{i,j}, \boldsymbol{q}_{i,j}) = \frac{\boldsymbol{s} \cdot [\boldsymbol{p}_{i,j}, \boldsymbol{q}_{i,j}, 1]^T}{\sqrt{1 + \boldsymbol{p}_{i,j}^2 + \boldsymbol{q}_{i,j}^2}} \tag{16}$$

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

$$\begin{aligned} 0 &= f(\boldsymbol{G}_{i,j}, \boldsymbol{Z}_{i,j}, \boldsymbol{Z}_{i-1,j}, \boldsymbol{Z}_{i,j-1}) \\ &= \boldsymbol{G}_{i,j} - \boldsymbol{R}(\boldsymbol{Z}_{i,j} - \boldsymbol{Z}_{i-1,j}, \boldsymbol{Z}_{i,j} - \boldsymbol{Z}_{i,j-1}) \end{aligned} \tag{17}$$

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

$$0 = f(\boldsymbol{Z}_{i,j}) \approx f(\boldsymbol{Z}_{i,j}^{n-1}) + (\boldsymbol{Z}_{i,j} - \boldsymbol{Z}_{i,j}^{n-1}) \frac{d}{d\boldsymbol{Z}_{i,j}} f(\boldsymbol{Z}_{i,j}^{n-1}) \tag{18}$$

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

$$\boldsymbol{Z}_{i,j}^{n} = \boldsymbol{Z}_{i,j}^{n-1} + \frac{-f(\boldsymbol{Z}_{i,j}^{n-1})}{\frac{d}{d\boldsymbol{Z}_{i,j}} f(\boldsymbol{Z}_{i,j}^{n-1})} \tag{19}$$

In our experiments, we use domain-specific average shape as the initial depth map. Figure 5 shows the steps of the proposed albedo estimation and shape recovery algorithm.

5. Experiments

In this section, we describe experiments performed to evaluate the usefulness and robustness of the albedo estimates obtained using the proposed image estimation framework. Figure 7 shows the albedo maps obtained from several images of a subject from the PIE dataset [14] taken under different illumination conditions. Average facial surface normals are used as $\boldsymbol{n}^{(0)}$. 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. As desired the estimated albedo maps appear quite similar to each other.

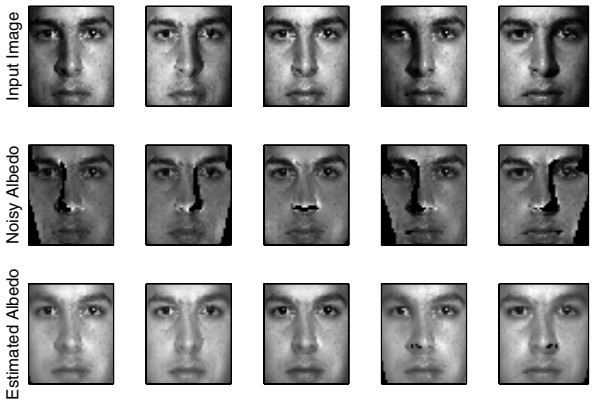

Figure 7. Albedo estimates obtained from several images of a subject from the PIE dataset [14].

We also perform a relighting experiment using the estimated albedo maps to generate images under frontal illumination. Figure 8 (second row) 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.

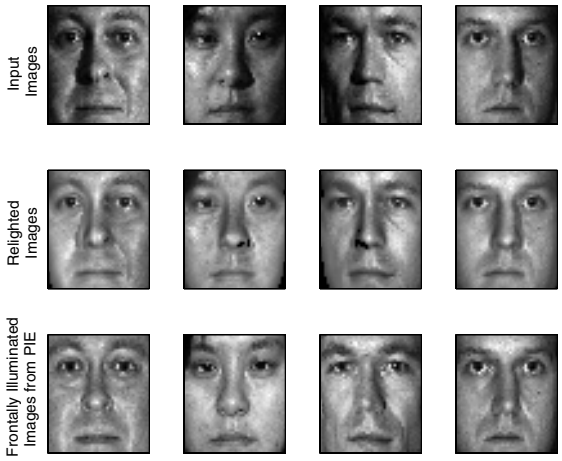

Figure 8. Relighting results on a few images from PIE [14].

5.1. 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,

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.

Table 1. Recognition results on the PIE dataset [14] using the estimated albedo. We include averages from [19] 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 [19]
Gallery														
f_{08}	-	100	100	99	93	91	79	72	44	100	96	85	87	89
f_{09}	100	-	100	100	99	97	91	90	75	100	99	93	95	93
f_{11}	100	100	-	100	100	97	88	78	57	100	100	93	92	92
f_{12}	99	99	100	-	100	100	96	96	87	100	100	97	98	96
f_{13}	99	99	100	100	-	100	99	99	90	99	100	100	99	98
f_{14}	97	99	100	100	100	-	99	97	90	100	100	100	98	99
f_{15}	84	94	88	100	100	100	-	100	99	93	100	100	96	96
f_{16}	76	97	79	99	100	99	99	-	100	75	99	100	93	91
f_{17}	53	82	56	90	96	94	94	100	-	54	96	97	83	80
f_{20}	100	100	100	100	100	100	94	78	57	-	100	99	93	91
f_{21}	99	99	100	100	100	100	93	94	85	100	-	97	97	96
f_{22}	90	99	97	100	100	100	100	97	91	97	100	-	97	98
Avg	91	97	93	99	99	98	94	91	80	93	99	96	94	-
Avg [19]	88	94	93	97	99	99	96	89	75	93	98	98	-	93

the similarity between images is measured using Principal Component Analysis (PCA). FRGC [11] 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 1 shows the rank-1 recognition results obtained.

The good recognition performance obtained using the estimated albedo maps justifies the usefulness of the proposed estimation framework. The performance compares favorably against the one obtained using the illumination-invariant approach in [19] that follows similar experimental setting. The performance is also comparable to the one reported by Romdhani *et al.* [12] and Zhang and Samaras [17]. 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. Initial albedo maps $\boldsymbol{\rho}^{(0)}$ 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. Note that the proposed estimation framework is able to take care of such large errors leading to good recognition performance.

5.2. Shape Recovery

We now demonstrate the usefulness of the albedo estimates obtained for the task of shape recovery. Figure 9 shows comparison of the recovered shapes with the ground truth on images from FRGC dataset [11]. The proposed approach seems to recover various person-specific facial features around lips, eyes, etc. Note that the ground truth shapes are captured on a different day than the input intensity images, leading to slightly different facial expressions in the estimated and true shapes.

We also test the efficacy of the approach on face images downloaded from the web (shown in Figure 10) with little control over the illumination and other imaging conditions. Figure 10 shows the albedo and shape estimates obtained along with a few novel views synthesized under novel illumination conditions.

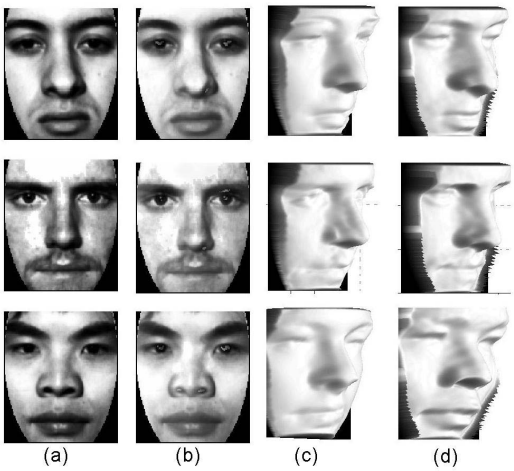

Figure 9. Comparison with the ground truth. (a) Input image, (b) Estimated albedo, (c) Recovered shape, (d) True shape

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.

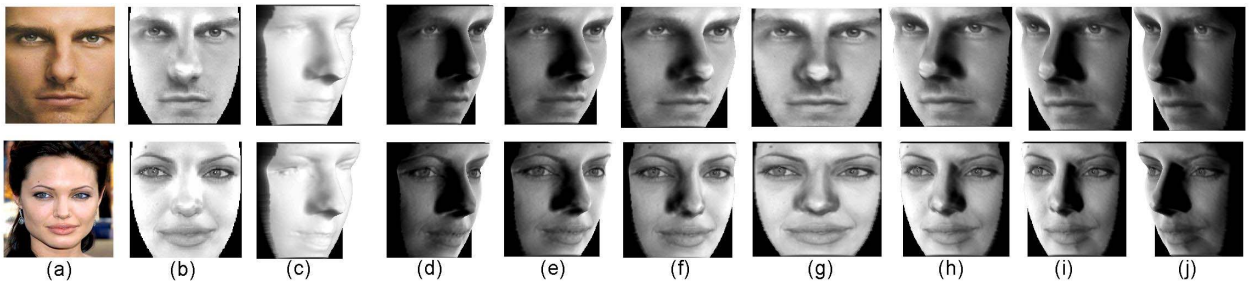

Figure 10. Novel view synthesis in the presence of novel illumination conditions. (a) Input image, (b) Estimated albedo, (c) Recovered shape, (d)-(j) Synthesized views under novel illumination conditions.

6. 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. 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 the task of shape recovery. The recovered shapes and albedo maps are used to synthesize novel views under novel illumination conditions. Though all our experiments are conducted on faces, the approach is applicable to any domain in general where the required error statistics can be estimated. In the absence of such information, ensemble statistics can be approximated using local spatial statistics [9]. The proposed algorithm is extensible to multiple light source scenarios as discussed in the supplemental material.

References

[1] G. L. Anderson and A. N. Netravali. Image restoration based on subjective criterion. *IEEE Trans. on Systems, Man and Cybernatics*, SMC-6:845–853, Dec 1976.

[2] H. C. Andrews and B. R. Hunt. *Digital Image Restoration*. Prentice-Hall signal processing series, 1977.

[3] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25:218–233, 2003.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, Sep 2003.

[5] M. J. Brooks and B. K. P. Horn. Shape and source from shading. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 932–936, Aug 1985.

[6] R. Dovgard and R. Basri. Statistical symmetric shape from shading for 3d structure recovery of faces. In *European Conference on Computer Vision*, 2004.

[7] B. K. P. Horn and M. J. Brooks. *Shape from Shading*. MIT Press, Cambridge Massachusetts, 1989.

[8] A. Hurlbert. Formal connections between lightness algorithms. *Journal of the Optical Society of America A*, 3(10):1684–1693, October 1986.

[9] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(2):165–177, March 1985.

[10] K. C. Lee and B. Moghaddam. A practical face relighting method for directional lighting normalization. In *International Workshop on Analysis and Modeling of Faces and Gestures*, Oct 2005.

[11] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[12] S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In *European Conference on Computer Vision*, pages 3–19, 2002.

[13] A. P. Sage and J. L. Melsa. *Estimation Theory with Applications to Comm. and Control*. McGraw-Hill, 1971.

[14] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[15] W. A. P. Smith and E. R. Hancock. Recovering facial shape using a statistical model of surface normal direction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):1914–1930, Dec 2006.

[16] P. S. Tsai and M. Shah. Shape from shading using linear approximation. *Image and Vision Computing Journal*, 12(8):487–498, 1994.

[17] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(3):351–363, March 2006.

[18] W. Zhao and R. Chellappa. Symmetric shape from shading using self-ratio image. *International Journal of Computer Vision*, 45(1):55–75, Oct 2001.

[19] S. Zhou, R. Chellappa, and D. Jacobs. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints. In *European Conference on Computer Vision*, 2004.

Authorized licensed use limited to: University of Maryland College Park. Downloaded on May 08,2010 at 04:37:57 UTC from IEEE Xplore. Restrictions apply.