Non-Lambertian Reflectance Modeling and Shape Recovery of Faces using Tensor Splines

Ritwik Kumar, *Student Member, IEEE,* Angelos Barmpoutis, *Member, IEEE,*
Arunava Banerjee, *Member, IEEE,* and Baba C. Vemuri, *Fellow, IEEE*

Abstract—Modeling illumination effects and pose variations of a face is of fundamental importance in the field of facial image analysis. Most of the conventional techniques that simultaneously address both of these problems work with the Lambertian assumption and thus, fall short of accurately capturing the complex intensity variation that the facial images exhibit, or recovering their 3D shape in presence of specularities and cast shadows. In this paper we present a novel Tensor Spline based framework for facial image analysis. We show that using this framework, the facial apparent BRDF field can be accurately estimated while seamlessly accounting for cast shadows and specularities. Further, using local neighborhood information, the same framework can be exploited to recover the 3D shape of the face (to handle pose variation). We quantitatively validate the accuracy of the Tensor Spline model using a more general model based on the mixture of single lobed spherical functions. We demonstrate the effectiveness of our technique by presenting extensive experimental results for face relighting, 3D shape recovery, and face recognition using the Extended Yale B and CMU PIE benchmark datasets.

Index Terms—Tensor Splines, Non-Lambertian Reflectance, Face Relighting, 3D Shape Recovery, Facial Image Analysis .

1 Introduction

Precisely capturing appearance and shape of objects has engaged human imagination ever since the conception of drawing and sculpting. With the invention of computers, a part of this interest was translated into the search for automated ways of accurately modeling and realistically rendering of appearances and shapes. Among all the objects explored via this medium, human faces have stood out for their obvious importance. In recent times, the immense interest in facial image analysis has been fueled by applications like face recognition (on account of recent world events), pose synthesis and face relighting (driven in part by the entertainment industry), among others. This in turn has led to tomes of literature on this subject, encompassing various techniques for modeling and rendering appearances and shapes of faces.

Our understanding of the process of image formation and the interaction of light and the facial surface has come a long way since we started [31], with many impressive strides along the way (e.g. [12], [23], [13]), but we are still some distance from an ideal solution. In our view, an ideal solution to the problem of modeling and rendering appearances and shapes of human faces should be able to generate extremely photo-realistic renderings of a person's face, given just one 2D image of the face, in any desired illumination condition and pose, at a click of a button (real time). Furthermore, such a system should not require any manual intervention and should not be fazed by the presence of common photo-effects like shadows and specularities in the input. Lastly, such an ideal system should not require expensive data collection tools and processes, e.g. 3D scanners, and should not assume availability of meta-information about the imaging environment (e.g. lighting directions, lighting wavelength etc.).

These general requirements have been singled out because the state-of-the-art is largely comprised of systems which relax one or more of these conditions while satisfying others. Common simplifying assumptions include applicability of the Lambertian reflectance model (e.g. [12]), availability of 3D face model (e.g. [9]), manual initialization (e.g. [7]), absence of cast shadows in input images(e.g. [10]), availability of large amounts of data obtained from custom built rigs (e.g. [23]) etc. These assumptions are noted as "simplifying" because – human faces are known to be neither exactly Lambertian nor convex (and thus can have cast shadows), fitting a 3D model requires time consuming large-scale optimization with manual selection of features for initialization, specialized data acquisition can be costly and in most realistic applications only a few images of a face are available.

The method we propose in this paper moves the state-of-the-art closer to the ideal solution by satisfying more of the above mentioned attributes simultaneously. Our technique can produce photo-realistic renderings of human faces across arbitrary illumination and pose using as few as 9 images (fixed pose, known illumination direction) with a spatially varying non-Lambertian reflectance model. Unlike most techniques, our method does not require input images to be free of cast shadows or specularities and can reproduce these in the novel renderings. It does not require any manual initialization and is a purely image based technique (no expensive 3D

All authors are with the Department of Computer and Information Science and Engineering at the University of Florida, Gainesville, FL 32611, USA

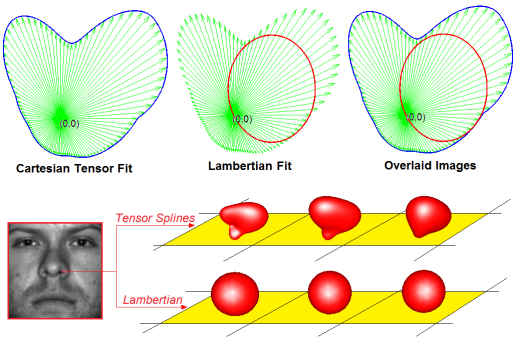

Fig. 1. **Lambertian Model vs. Cartesian Tensors**. From the synthetic example of the first row and the real data below it, it can be noted that Cartesian tensor can capture variations of intensity distributions more accurately than the Lambertian model.

scans are needed). Furthermore, it is capable of working with images obtained from standard benchmark datasets and does not require specialized data acquisition.

Our technique is based on a novel framework of Tensor Splines which can be used to approximate any n-dimensional field of spherical functions. In the case of faces, we use Tensor Splines to approximate the field of Apparent Bidirectional Reflectance Distribution functions (ABRDF) for a fixed viewing direction. Unlike the BRDF, the ABRDF (also known as the Reflectance Field) at each pixel captures the variation in intensity as a function of illumination and viewing direction and is thus sensitive to the context of the pixel. Once the ABRDF field has been captured, images of the face under the same pose but with arbitrary illumination can be generated by simply taking weighted combinations of the ABRDF field samples. Next, we estimate the surface normal at each pixel by robustly combining the shape information from its neighboring pixels. To this end, we put forth an iterative algorithm which works by registering neighboring ABRDFs using an extremely efficient linear technique. With as few as 1 or 2 iteration, we can recover the surface normal fields of most faces which are then numerically integrated to obtain the face surfaces. Novel poses with novel illumination conditions can then be rendered while seamlessly accounting for attached as well as cast shadows. *Note that in this paper we use the term "ABRDF" to refer to a spherical function of illumination direction with the pose held fixed.*

2 Related Work

The sheer size of the facial shape-reflectance modeling literature allows its taxonomy to be carried along various lines. Since the work presented in this paper deals with a novel ABRDF model and subsequent shape recovery, we would categorized the methods based on the assumed reflectance models (Lambertian or Non-Lambertian). We have focused primarily on bringing out the underlying assumptions and limitations of the methods with respect to the problem of relighting and shape recovery. In particular, we are interested in knowing the number of required input images, use of external 3D information and handling of cast shadows by different methods. Note that some of the surveyed methods deals with objects in general, but are included here because they can be easily adapted to work with faces. We have summarized few of the key methods along with the associated assumptions in Table 1.

A large fraction of the existing techniques for facial image analysis work with the Lambertian assumption for reflectance. This translates to assuming that the apparent BRDF at each point on the object's surface has the same shape, that of a half cosine function, which has been scaled by a constant – the albedo, and is oriented along the surface normal at that location. One of the major reason for the prevalence of this model is its simplicity. Analysis has shown that under this assumption, if cast and attached shadows are ignored, image of a convex object, in a fixed pose, lit by arbitrary illumination lies in a 3-dimensional subspace [32]. When an ambient lighting component is included, this subspace expands to become 4-dimensional [33] and when attached shadows are taken into account, the subspace grows to become an infinite dimensional – illumination cone [34]. Note that the Lambertian model does account for attached shadows but not cast shadows.

Spherical harmonic analysis of the Lambertian kernel has shown that even though the illumination cone is infinite dimensional, it can be approximated quite well by a lower dimensional subspaces ([36], [13], [12]). In particular, these methods can produce impressive results with 9 basis images, though they require 3D shape and the albedo field as input. These basis images can also be directly acquired using the "universal virtual" lighting conditions [40]. More recently, this idea has been extended to 3D surfaces in [7] building on the prior seminal work presented in [9] called Morphable Models. Morphable Models can recover 3D shape of a face by fitting an average 3D facial model to a given 2D image, accounting for necessary shape and texture adjustments. Morphable Models are known to produce excellent results for across pose face recognition but cannot handle cast shadows or specularities robustly. More importantly, they require manual delineation of facial features to initialize a complicated non-linear optimization which can take a long time to converge and can suffer from local minima. Use of a single reference 3D face to recover shape from a single image was explored in [55]. Using the idea of low dimensional subspace mentioned above, [30] represented the entire light-field using a low dimensional eigen light-field.

It has been suggested that even though the time and cost of acquiring the 3D data is decreasing, the majority of the face databases still remain 2D and hence it is more pragmatic to work with 2D images alone [22]. Methods that are purely image based and work with the Lambertian assumption generally apply photometric

TABLE 1
Requirements, Assumptions and Capabilities of Candidate Methods for Relighting and/or Pose Change

Methods	Assumed Surface BRDF Model	No. of Images as Input	Relit Images Presented	Shape or Pose Results Presented	Cast Shadows in Input	Purely Image based (No 3D Scans)	Other Assumptions, Requirements and Limitations
1999 ***Georghiades*** **et al.** [19] ☺	Lambertian	≥ 3	✔	✔	✘	✔	Near frontal illumination expected, Ray tracing for cast shadows.
1999 ***Blanz*** **et al.** [9] ☺	Non-Lamb.	1	✔	✔	✘	✘	No attached shadows, Manual initialization to fit 3D model.
2000 ***Debevec*** **et al.** [23] ☺	Non-Lamb.	≥ 2000	✔	✔	✔	✔	Custom rig for data collection, Structured lighting for shape.
2001 ***Ramamoorthi*** **et al.** [13]	Lambertian	≥ 3	✔	✘	✘	✘	Distant and isotropic lighting, 3D Scans needed as input.
2001 ***Malzbender*** **et al.** [35]	Non-Lamb.	≥ 50	✔	✔	✘	✘	Custom rig for data acquisition, No specularity allowed in input.
2001 ***Georghiades*** **et al.** [10] ☺	Lambertian	≥ 7	✔	✔	✘	✔	Almost no attached shadow, Symmetric faces, Ray tracing.
2001 ***Shashua*** **et al.** [29] ☺	Lambertian	1	✔	✘	✘	✔	Bootstrap set of images required, Ideal class assumption.
2001 ***Zhao*** **et al.** [25] ☺	Lambertian	1	✔	✔	✘	✔	No attached shadows, Symmetric faces, piecewise constant albedo.
2001 ***Magda*** **et al.** [24]	Non-Lamb.	≥ 300	✘	✔	✔	✔	Known lighting directions, Lighting should doubly cover the directions
2003 ***Georghiades*** [20] ☺	Non-Lamb.	≥ 12	✔	✔	✘	✔	3 sources/pixel, ad-hoc shadow detection, Spatially constant BRDF.
2003 ***Wen*** **et al.** [3] ☺	Lambertian	1	✔	✔	✘	✘	Symmetric lighting, 3D Model Fitting, Manual initialization.
2003 ***Basri*** **et al.** [12] ☺	Lambertian	1	✔	✘	✘	✘	Distant & isotropic lighting, 3D scans required, Manual initialization.
2004 ***Gross*** **et al.** [30] ☺	Lambertian	≥ 1	✘	✔	✘	✔	Manual delineation of feature points for better recognition.
2005 ***Goldman*** **et al.** [21]	Non-Lamb.	12	✔	✔	✘	✔	Known lighting, HDR images expected, Manual threshold selection
2005 ***Hertzman*** **et al.** [18]	Non-Lamb.	≥ 8	✘	✔	✘	✔	No shadows expected, Reference object expected, Symmetry of faces.
2005 ***Lee*** **et al.** [28] ☺	Lambertian	1	✔	✘	✔	✘	Shadowed pixel gets default albedo, Universal 3D face model required.
2006 ***Zhang*** **et al.** [7] ☺	Lambertian	1	✔	✔	✘	✘	3D Model Fitting with manual initialization.
2006 ***Zickler*** **et al.** [14] ☺	Non-Lamb.	≥ 1	✔	✘	✘	✘	Point lighting sources with known directions, Object shape required.
2007 ***Chandraker*** **et al.** [4]	Lambertian	≥ 4	✘	✔	✔	✔	3 sources/pixel, Known lighting, Normals can't be on bisection planes.
2007 ***Alldrin*** **et al.** [15]	Non-Lamb.	≥ 32	✘	✔	✔	✔	Point light sources, Known directions, BRDF isotropic about normal.
2007 ***Biswas*** **et al.** [27] ☺	Lambertian	1	✔	✔	✘	✘	Point sources with known directions, Registered avg. 3D model required.
2007 ***Zhou*** **et al.** [5] ☺	Lambertian	1	✔	✔	✘	✔	No attached shadows expected, Symmetry of faces, Bootstrap set required.
2007 ***Basri*** **et al.** [11]	Lambertian	15	✘	✔	✘	✔	Distant and isotropic lighting, Works for only convex objects.
2008 ***Alldrin*** **et al.** [16]	Non-Lamb.	≥ 102	✔	✔	✘	✔	Point sources with known directions, BRDF isotropic about normal.
Our Method ☺	Non-Lamb.	≥ 9	✔	✔	✔	✔	Point sources with known directions.

☺ *indicates that the method was primarily demonstrated on face images.*

stereo or shape from shading to recover facial shape from the given images. For instance, results for simultaneous shape recovery using photometric stereo and reflectance modeling were presented in [19] and [10]. Both of these methods work with multiple images and expect no cast shadows and very little attached shadows in the images. Here the cast shadows in the relit images are rendered using ray tracing, which can be computationally expensive. Examples of methods that recover shape from shading working under the Lambertian assumption can be found in [25] and [38]. As these methods work with just a single image, besides requiring absence of cast shadows in input, they make additional assumptions like facial symmetry, as in [25]. An important point to note here is that *uncalibrated* photometric stereo or shape from shading methods that work with the Lambertian assumption and orthographically projected images, also suffer with the Bas-Relief Ambiguity ([37]). Resolving this requires additional assumptions like symmetry of face, nose and forehead being at the same height, known lighting directions, etc., and manual assistance.

Recently, shape recovery using generalized photometric stereo, for objects in general, was presented in [11] which relaxes some of the assumptions made by traditional photometric stereo. This method can recover shape from images taken under general unknown lighting. On account of the Lambertian assumption, cast shadows are not entertained in the input images and the shape of the object is assumed to be convex. Note that accurate recovery of shape using this method requires 15 to 60 images as input. Another method for Lambertian shape recovery with multiple illuminants, but without ignoring shadows, was presented in [4] where the graph cuts method was used to identify light source visibility and information from shadow maps were used to recover the shape. This method is again meant for objects in general and but can be applied to faces.

In contrast to most of the methods mentioned above are the techniques that seek illumination invariant representations of faces which can be then used to render relit images. Seminal work in this category was presented in [29], where the so called "Quotient Images", generated using ratio of albedo values, were used to generate images under novel illumination. More recently, use of invariants was invoked in [3], where the radiance environment map was deduced using the ratio image technique ([39], [29]). Note that the shape recovery in [3], like the Morphable Models, requires manual initialization. Forgoing the ratio technique, direct use of albedo as a illumination invariant signature of face images was explored in [28], where using an universal 3D face model, illumination normalized images of faces were generated. This method worked with low resolution images and did not render high quality relit images. More recently an improvement was presented in [27] where the albedo estimation was made more robust using the error statistics of surface normals and the known illumination direction. This method requires a registered average 3D model of the face and does not allow cast shadows in the input but as compared to [28], it provides better shape recovery. Improving upon the idea of ideal class assumption ([29]), a generalized photometric stereo was presented in [5]. Using a bootstrap set of facial images and exploiting the subspace spanned by a set of basis objects with Lambertian surfaces, images with novel pose and illumination were generated. Faces were assumed to be symmetric and the input was assumed to be free of shadows.

Next, we look at techniques that do not make the Lambertian assumption. Seminal work in this class of techniques was presented in [23] where using a custom built rig, dense sampling of the illumination space for faces was obtained. In this work the facial shape was obtained using structured lighting and no assumption about the surface BRDF was made. This completely data driven technique was able to produce extremely photo-realistic images of the face in novel illuminations and poses. The specular component was captured using polarized lighting and modified appropriately for pose variation. This method demonstrated that if a large number of images (> 2000) for each subject can be obtained under various lighting configurations, the relighting and pose generation problem can be solved, but the cost of such a system can be extremely high.

Use of biquadratic polynomials to model texture was explored in [35]. This method required a custom built rig and more than 50 specularity free images to recover the model parameters. The shape of the object was not recovered in this method. Use of a large number (≥ 300) of images to recover the shape without making any assumption about the nature of the BRDF was revisited in [24]. This method required input images to doubly cover the illumination direction which called for specialized data acquisition. Though no attempt to capture the reflectance properties of the object was made in this work, it was addressed in the follow-up work presented in [53]. [53] was also possibly the first paper to coin the phrase "Apparent BRDF". Note that all these methods were targeted at general objects but can be applied to faces.

One of the first non-Lambertian techniques that worked with standard face databases and did not require custom data was presented in [20] where the more general Torrance-Sparrow ([47]) model for BRDF was used. This method presented relighting and pose variation results with 12 images as input but did not allow cast shadows. Further, this method required each pixel to be lit by at least 3 light sources in order to work properly.

Important contribution in the field of example based shape recovery was made by [17] where objects of interest were imaged along with a reference object of known geometry (e.g. sphere). Multiple (≥ 8) cast shadow free images were used as input. This work was expanded to allow spatially varying BRDF by using multiple references in [18]. An interesting extension of this work was presented in [21] where the shape of an object was recovered using the assumption that the BRDF of any object is essentially composed of BRDFs of a few fundamental materials. This method used 12 High Dynamic Range images with known illumination direction and required manual selection of a system parameter threshold. These methods work with general objects and can also be used with faces. Recently [57] presented a method where 16 to 59 input images are used recover the shape and the reflectance of objects. Though useful in certain cases, this method suffers from large memory, close initialization and camera calibration requirements.

When the 3D shapes of the objects are assumed available, [14] presented a technique which, at times using just 1 image, can recover their spatially varying non-parametric BRDF fields. For the case of the human face, this work presented results with 4 images where specular component was separately captured using polarized lighting. The images were acquired from known illumination directions and no cast shadows were allowed.

Recently, [15] presented a new method for photometric reconstruction of shape, which can be applied to faces, assuming spatially varying but isotropic BRDFs. Given 32 or more images with known illumination, this method recovers isocontours of the surface depth map from which shape can be recovered by imposing additional constraints. An extension of this work was presented in [16] where the need for additional constraints to recover shape from the depth map isocontour was alleviated by assuming the surface to be composed of a few fundamental materials and that the BRDF at each point can be approximated by a bivariate function. Results presented in this work required 102 or more images. Another interesting framework for photometric stereo using the Markov Random Field approach was presented in [51].

Lastly, we note that in cases when extremely high quality renderings are required and cost-time constraints are relaxed, custom hardware is employed. For instance, highly accurate measurements of material BRDF were carried out using a gonioreflectometer in [42], various customized hardware components and software were used to render face images in the movie "The Matrix Reloaded" [41]. In order to measure accurate skin reflectance while accounting for sub-surface scattering, custom built devices were again employed in [43] to render high quality facial images.

It can be noted that most of the image based techniques that do not make the simplifying Lambertian assumption end up using a large amount of custom acquired data or assuming some other parametric form for BRDF (besides the other assumptions). In this paper we explore the possibility of acquiring the non-Lambertian reflectance and shape with just nine images in a purely data driven fashion.

3 Overview

The technique that we propose in this paper simultaneously captures both shape and reflectance properties of a face. Unlike the majority of existing techniques that work with BRDFs, in order to seamlessly account for specularities, attached shadows, cast shadows and other photo-effects, we have chosen to work with the ABRDFs, which are spherical functions of non-trivial shape. We estimate them using Cartesian tensors, which in practice, have enough flexibility to account for the variations in ABRDF across the human face. Further, in order to robustly estimate the ABRDF field from only a few and often noisy samples, we draw upon the apparent

smooth variation of reflectance properties across the face and combine the Cartesian tensors with B-Splines. This combination of Cartesian tensors with B-Splines is called Tensor Splines in this paper.

Embedded in the ABRDFs at each pixel also lies the surface normal of the shape. To extract the normal from the ABRDF field riddled with cast shadows and specularities, we invoke the homogeneity of the ABRDFs in local neighborhoods, and infer surface normal at a pixel using the information from its immediate neighbors. More concretely, at each pixel we align the ABRDF with its neighbors using a linearized algorithm for rotation recovery and take a weighted geodesic mean of the normals suggested by the neighbors to obtain the surface normal. Our framework automatically discounts possibly erroneous surface normal suggestions by weighting the suggestion from a neighbor of substantially different ABRDF shape lower than others. This process can be iterated and in practice we find good solutions within 1 or 2 iterations.

Equipped with this mechanism to capture both reflectance properties and shapes of the human faces, we can generate images of any face in novel poses and illumination conditions.

3.1 Assumptions

Like all other techniques, our method also works with certain assumptions. It requires at least 9 images of the face under point illuminations from known directions in a fixed pose. Note that these assumptions have been used in the past by various methods, for example, [21] worked with 12 images obtained from known lighting directions in fixed pose. As the number of input images increases so does the performance of our method. We do not restrict input images to be free of attached or cast shadows. We also do not restrict the BRDF to be Lambertian ([10]) or isotropic ([15], [14]). Though global photo-effects like subsurface scattering and interreflection are not explicitly modeled, Tensor Splines can capture them to some extent.

4 Tensor Splines

We seek a mathematical framework that can represent a field of spherical functions accurately. If a dense enough sampling of the spherical function field is provided, this can be accomplished to arbitrary accuracy, but the central problem we face is precisely the scarcity of the data. To solve this problem for the case of human facial ABRDF fields, we exploit clues from the specific nature of ABRDFs on human faces e.g. smooth variation of ABRDF for the most part, presence of multiple lobes in the ABRDF etc.

4.1 Spherical functions modeled as Tensors

A spherical function in $\mathbb{R}^3$ can be thought of as a function of directions or unit vectors, $\mathbf{v} = (v_1\ v_2\ v_3)^T$. Such a function, T, when approximated using an n^{th} order [1] Cartesian tensor [56] (a tensor in $\mathbb{R}^3$), is expressed as

$$T(\mathbf{v}) = \sum_{k+l+m=n} T_{klm}(v_1)^k(v_2)^l(v_3)^m \quad (1)$$

where T_{klm} are the real-valued tensor coefficients and k, l & m are non-negative integers. This is a Cartesian tensor with all the n arguments set to be $\mathbf{v}$. The expressive power of such Cartesian tensors increases with their order. Geometrically this translates to presence of more "lobes" on a higher order Cartesian tensor.

Note that the Lambertian model is intricately connected to a special case of the Cartesian tensor formulation. If $\mathbf{v} = (v_1\ v_2\ v_3)^T$ is the light source direction, $\mathbf{n} = (n_1\ n_2\ n_3)^T$ is the surface normal and ρ is the surface albedo, the Lambertian kernel is given by

$$\begin{aligned} max(\rho \cdot \mathbf{n} \cdot \mathbf{v}, 0) &= \rho \cdot max(n_1v_1 + n_2v_2 + n_3v_3, 0) \\ &= max(\sum_{k+l+m=1} T_{klm}v_1^k v_2^l v_3^m, 0) \quad (2) \end{aligned}$$

with $T_{100} = \rho \cdot n_1$, $T_{010} = \rho \cdot n_2$ and $T_{001} = \rho \cdot n_3$. A comparison with Eq. 1 reveals that the Lambertian kernel is exactly the positive half of the 1^{st} order Cartesian tensor.

The 1^{st}, 2^{nd}, 3^{rd} and 5^{th} order Cartesian tensors have 3, 6, 10 and 21 unique coefficients respectively. For even orders, the Cartesian tensors are symmetric, $T(\mathbf{v}) = T(-\mathbf{v})$, while for odd orders they are anti-symmetric, $T(\mathbf{v}) = -T(-\mathbf{v})$. We must point out that these definitions of symmetry and anti-symmetry are different than the standard definition based on switching of the arguments' order. In this paper, we would use the definitions we provided above.

4.2 Tensor Splines

When the task requires estimation of a p-dimensional field of multi-lobed spherical functions from sparse and noisy data, given the high noise sensitivity of higher order tensors, it is reasonable to enforce smoothness across the field of spherical functions. We accomplish this by combining the Cartesian tensor basis at each pixel with the B-Spline basis ([44]) across the lattice of spherical functions.

We define a Tensor Spline as a B-spline of multilinear functions of any order. In a Tensor Spline, the multilinear functions are weighted by the B-spline basis $N_{i,k+1}(t)$ [52]. The $N_{i,k+1}(t)$ are polynomials of degree k, associated with n+k+2 monotonically increasing numbers called "knots" $(t_{-k}, t_{-k+1}, ..., t_{n+1})$ and i is the index of the control points.

The Tensor Spline for a p-dimensional lattice of spherical functions, with k^{th} degree spline and n^{th} order Cartesian tensor is defined as

$$S(\mathbf{t}, \mathbf{v}) = \sum_{(i_1 \ldots i_p) \in \mathfrak{D}} (\prod_{i_a} N_{i_a,k+1}(t_{i_a}))T_{i_1 \ldots i_p}(\mathbf{v}) \quad (3)$$

1. In the notation used in this paper, this order is not same as the number of indices used to represent the tensor.

where $\mathbf{t} = (t_1 \dots t_p)$ is the index into the spherical function lattice, $\mathbf{v} = (v_1\ v_2\ v_3)^T$ is a unit vector, $\mathfrak{D}$ is the p-dimensional spline control point lattice and $T_{i_1 \dots i_p}(\mathbf{v})$ is given by Eq. 1. In Tensor Splines the usual B-Spline control points have been replaced by control tensors $T_{i_1 \dots i_p}(\mathbf{v})$. The formulation presented in Eq. 3 is quite general as it can be used to estimate a spherical function field defined over an arbitrary dimensional lattice, with the desired degree of B-Spline smoothing.

4.3 Facial ABRDF approximation using Tensor Splines

Human faces are known to be neither exactly Lambertian nor convex, which leads to photo-effects like specularities (oily forehead and nose tip) and cast shadows (around protruding features like nose and lips) in facial images. These effects cause such a complex variation in the intensity values at various pixels as the lighting direction changes that it cannot be captured by a single lobed function (like the Lambertian kernel). This motivated us to explore the use of higher order Tensor Splines to model the ABRDFs. Note that here the lattice is 2-dimensional and the assumption of local homogeneity also holds to a reasonable degree in case of facial ABRDFs. In order to ensure that the smoothness is manifested only in a localized fashion, we have chosen to use bi-cubic B-Splines in the ABRDF-specialized version of Tensor Splines. This smoothing may cause the estimated ABRDFs to be only an approximation of the true functions at the material boundaries, but since the skin reflectance is largely uniform across the face, this approximation does not degrade rendered image quality.

The ability of the Cartesian tensors to better model data with complex distributions can be noted in Fig. 1, where in the first row we show that for the case of synthetic circular data (shown by green arrows), the Cartesian tensors can more accurately approximate the data than the Lambertian cosine bumps. In the second row we show real facial ABRDFs approximated by the Tensor Splines and the Lambertian model from a shadow prone region of the face. It can be readily noted that the Tensor Splines capture the variability in intensity values, as a function of illumination direction, more accurately than the Lambertian reflectance model.

We must point out that as the order of Cartesian tensors increases, so does the amount of data samples required to estimate the unknown coefficients. When there are only a few images available, in order to satisfy our desire to use higher order tensors, we must choose between its odd (anti-symmetric) or even (symmetric) components. Note that since most of the time we are interested in the ABRDFs' behavior on the frontal hemisphere, both symmetric and anti-symmetric versions provide the same representation power. Their behavior only becomes pertinent when the illumination direction is exactly perpendicular to the pose direction, and this is where the use of anti-symmetric versions is advantageous.

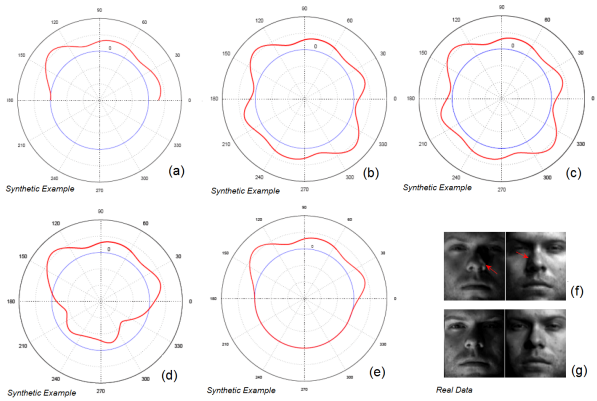

Fig. 2. Symmetric and Antisymmetric ABRDF approximations.

This has been explained via a 2D example in Fig. 2 where case (a) shows a semicircular function. The blue circle in the figure is considered to be the zero value. Cases (b) and (d) show the same function being approximated by a antipodally symmetric and anti-symmetric functions respectively. In can be noted that the approximation is quite accurate except near the angles $0°$ and $180°$. When the original function (case (a)) is such that it has positive value at one of these antipodal points and near zero value at the other, a symmetric function forces value at both of these crucial angles to be positive while the anti-symmetric function force one to be positive and other to be negative. Now, if we assume that only the positive values of the function are preserved we get the results as presented in cases (c) and (e).

The behavior of most facial ABRDFs is similar to the function in case (a). This is because if a pixel has high intensity value when lit from $0°$, most of the time it would have a low intensity value when lit from $180°$ (due to attached and cast shadows), and vice versa. Thus, if a symmetric function is used for approximating such an ABRDF, it would cause non-negative values at both $0°$ and $180°$ and would lead to visually significant artifacts (unnatural lighting) in the images (case (f)). On the other hand, in practise, use of an anti-symmetric function does not cause visually significant artifacts (case (g)). To summarize, even though both, anti-symmetric and symmetric functions, introduce artifacts near $0°$ and $180°$ directions, the artifacts created by an anti-symmetric approximation are visual insignificant and hence we have chosen to work with anti-symmetric components.

Two dimensional Tensor Splines with bi-cubic B-Splines and odd order tensors can be written as

$$S(\mathbf{t}, \mathbf{v}) = \sum_{(i,j) \in \mathfrak{D}} N_{i,4}(t_x) N_{j,4}(t_y) T_{i,j}(\mathbf{v}) \qquad (4)$$

where vectors $i, j, \mathfrak{D}, \mathbf{t}$ and $\mathbf{v}$ have the same meaning as before and the tensor has an odd order.

The problem at hand is that given a set of Q face images ($I_q,\ q = 1 \dots Q$) of a subject in a fixed pose along with associated lighting directions $\mathbf{v}_q = (v_{q1}\ v_{q2}$

v_{q3}), we want to estimate the ABRDF field of the face using a bi-cubic Tensor Spline. We propose to accomplish this by minimizing the following energy function which minimizes the L_2 distance between the model and the given data,

$$\mathbf{E}(T_{ijklm}) = \sum_{q=1}^{Q} \sum_{t_x,t_y} \Big(\sum_{(i,j)\in\mathfrak{D}} N_{i,4}(t_x) N_{j,4}(t_y) T_{i,j}(\mathbf{v}_q) - I_q(t_x,t_y) \Big)^2 = \sum_{q=1}^{Q} \sum_{t_x,t_y} \Big(\sum_{(i,j)\in\mathfrak{D}} N_{i,4}(t_x) N_{j,4}(t_y) \sum_{k+l+m=n} T_{ijklm} v_{q1}^k v_{q2}^l v_{q3}^m - I_q(t_x,t_y) \Big)^2 \quad (5)$$

where t_x, t_y run through the lattice of the given images, i, j are the indices into the spline control point lattice $\mathfrak{D}(D \times D)$, and the tensor order n is an odd integer. The minimization of Eq. 5 is done with respect to the unknown tensor coefficients $T_{i,j,k,l,m}$ that correspond to the control tensors $T_{i,j}(\mathbf{v}_n)$.

If the image size is $M \times M$, there are M^2 unknown ABRDF tensors which are interpolated from the control tensors (Eq. 4). We use a uniform grid $D \times D$ of control tensors, which translates to $3D^2$, $10D^2$ and $21D^2$ unknown control tensor coefficients for 1^{st}, 3^{rd} and 5^{th} order tensors respectively. A value for D is chosen according to the desired smoothness. For the cases when the number of unknowns per control tensor is one more than the number of data constraints, we use an additional constraint which discourages solutions with large norms. This is enforced by adding the term $\lambda \sum_{ij} \sum_{klm} T_{ijklm}^2$ to the error function in Eq. 5, where λ is the regularization constant.

We recover the unknowns in Eq. 5 using the gradient descent method with the control tensor coefficient field initialized using all-ones unit vectors. This technique can be efficiently implemented because the closed form for the derivative of the objective function with respect to the unknown coefficients can be easily obtained analytically.

Once the coefficients have been recovered, images under novel illumination direction, $\mathbf{v}$, can be synthesized by evaluating the ABRDF field in the direction $\mathbf{v}$, where each ABRDF is given by Eq. 3. Possible negative values obtained in Eq. 3 are set to zero (as in Lambertian model). Furthermore, it should be noted that the generated images can be readily up-sampled by evaluating Eq. 3 on a more dense sampling lattice since the Tensor Spline is a continuous function.

5 Mixture of Single-Lobed Functions

In order to quantitatively validate whether the Tensor Splines provide a good enough approximation of the ABRDF field, we present a more expressive model here. This validation model is more general in the sense that it can accommodate arbitrarily large number of lobes to approximate any spherical function. We define it using

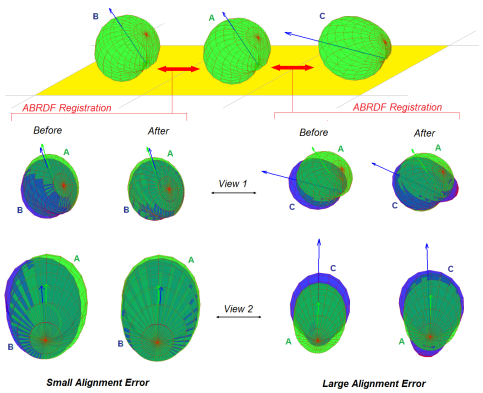

Fig. 3. **ABRDF alignment**. Neighboring ABRDFs A and B can be better aligned with each other than A and C. This would weigh the normal suggested by B higher than the normal suggested by C.

a mixture of single-lobed spherical functions. Such a mixture is characterized by a kernel function, $k(\mu_i, \mathbf{v})$, and a set of mixing weights, w_i, associated with a set of unit vectors μ_i as follows

$$B(\mathbf{v}) = \sum_i w_i k(\mu_i, \mathbf{v}), \quad (6)$$

where $\mathbf{v}$ is the lighting direction and the vectors $\mu_\mathbf{i}$ are uniformly distributed on the unit sphere.

Of the various choices for singled lobed spherical functions that can be used as the kernel function $k(\mu, \mathbf{v})$, we picked $k(\mu, \mathbf{v}) = e^{-\mu \cdot \mathbf{v}} - 1$ due to two reasons – it has a single peak and $k(\mu, \mathbf{v}) = 0$ for all $\mathbf{v}$ such that $\mathbf{v} \cdot \mu = 0$ (since if the viewing and illumination directions are perpendicular we expect zero intensity). Note that the these two properties are also satisfied by the Lambertian kernel.

The task of estimating ABRDFs using this mixture model requires us to recover the unknown weights such that the weighted combination leads to a spherical function which closely approximates the ABRDFs. Given a set of N facial images with the same fixed pose with associated lighting directions $\mathbf{v}_n$, we can setup a $N \times M$ matrix $\mathbf{A}_{n,m}$ by evaluating $e^{-\mu \cdot \mathbf{v}} - 1$ for every $\mathbf{v}_n$ and μ_i. M is the number of μ_i picked in the model. The unknown weights (Eq. 6) for each pixel can then be estimated by solving the overdetermined system $\mathbf{AW} = \mathbf{B}$, where $\mathbf{B}$ is an N-dimensional vector of the intensities at a fixed pixel in the N given images, and $\mathbf{W}$ is the vector of the unknown weights. Since ABRDF is a nonnegative function, we solve this system with the positivity constraint using the non-negative least square minimization algorithm developed in [45].

Note that this model would generally have a very large number of unknowns (depending on the chosen resolution while picking μ_i), and thus would require a large number of ABRDF field samples (images) for accurate recovery of the ABRDFs. But since this would only be used as a tool to evaluate the Tensor Splines, it is not considered a drawback.

6 Recovering Shape from the ABRDF field

Facial ABRDF is in part characterized by the local surface normal and hence it should be possible to recover shape information from it. But unlike the various popular parametric reflectance models like Lambertian, Torrance-Sparrow ([47]), Phong([46]) etc., which explicitly assume a role for the surface normal in their formulae, Tensor Splines make no such assumption. This allows spatially varying and accurate approximation of the ABRDFs, but also makes the recovery of the surface normal non-trivial.

To recover the surface normal from the Tensor Spline model we invoke the local homogeneity of the ABRDF field. This assumption is physically sound because the reflectance properties of a human face do not change drastically in small neighborhoods (3×3 pixels) and mathematically robust as Tensor Splines ensure that the coefficients vary smoothly across the ABRDF lattice. We assume that ABRDFs at two neighboring pixels have the same shape and differ only by a rotation, R and thus, if the surface normal at one of these pixels is known, the surface normal at the other pixel can be derived by rotating it by R.

For a given internal pixel (x, y) in the image, there are eight immediate neighbors. If the surface normal at (x, y) is inferred as described above, it would receive eight suggestions for possible surface normals (assuming that the surface normals for the neighbors are known). Instead of picking one of the suggestion as its surface normal, we take a weighted geodesic average of the suggested vectors. The weights are set to be inversely proportional to the registration error obtained during rotation-alignment of the ABRDF pairs. There are two main advantages to computing the surface normal is this manner. Firstly, being an aggregate statistic, the geodesic mean is more robust to noise than the individual suggestions. Secondly and more importantly, the weighted nature of the mean ensures that suggestions, which originate from neighbors whose ABRDFs are very different in shape than the ABRDF at (x, y), are automatically weighted less. This property of the mean is specially useful at locations in the image where the homogeneity assumption breaks down, e.g. shadow edges.

This processes is summarized in Fig. 3, where the central ABRDF (A) is shown to be aligned with a higher accuracy to its left neighbor ARBDF (B) than to its right neighbor ABRDF (C). For both case, before and after alignment configurations are shown from 2 different points of view. As mentioned before, the misalignment error is used to weight the normal suggestion from a neighbor and hence the suggestion from the left ABRDF (B) would eventually be weighted more than the suggestion from the ABRDF on the right (C).

Once the rotation matrices for all the pixels in the image have been computed, we initialize all the normals with the directions in which ABRDFs have their maxima. Initialization is followed by weighted geodesic mean computations which provides us with a robust estimate of the surface normals. The process of mean computation is carried out iteratively but empirically it was noticed that good results can be obtained in all cases with 1 or 2 iterations. Note that using the maxima directly as a normal estimate provides inaccurate results. We attribute this to the fact that unlike some reflectance models (e.g. Lambertian), Tensor Spline do not ensure that the maximal response of the ABRDF lies along the surface normal direction.

6.1 Rotation Estimation

Recovering the surface normal field using the steps described above requires computation of rotation matrices for each pair of neighboring ABRDFs in the image. A simple but computationally intensive approach would be to search for the rotation matrix using a gradient based constrained optimization technique. More concretely, two ABRDFs, represented by their Cartesian tensor coefficients w_1 and w_2, can be aligned by minimizing the following objective function

$$E(R) = \sum_{\mathbf{v} \in S^2} (w_1^T B(\mathbf{v}) - w_2^T B(R \cdot \mathbf{v}))^2. \tag{7}$$

such that

$$R^T R = I. \tag{8}$$

where unit vector $\mathbf{v}$ is obtained by some uniform sampling of directions on a sphere (e.g. 4th-order tessellation of a icosahedron provides 642 fairly uniformly distributed directions), B is the vector of Cartesian tensor basis defined in Eq. 1 and R is the sought rotation matrix. This method for rotation matrix recovery would require nonlinear optimization to be run $\sim 8L^2$ times for an image of size $L \times L$ pixels. Even for an average sized image this process can be quite intractable and hence, we propose the following more efficient algorithm for the rotation matrix recovery.

Let $T_1(\mathbf{v})$ and $T_2(\mathbf{v})$ be the two ABRDFs (Eq. 1) that need to be aligned via a rotation. This implies that for each $\mathbf{v}$, we seek a $\delta\mathbf{v}$ such that

$$T_1(\mathbf{v}) = T_2(\mathbf{v} + \delta\mathbf{v}). \tag{9}$$

Since the ABRDFs are from neighboring pixels, we assume that the required $\delta\mathbf{v}$ would be small and thus using the first order Taylor's expansion, we get

$$T_1(\mathbf{v}) = T_2(\mathbf{v}) + \nabla T_2(\mathbf{v})^T \delta\mathbf{v}. \tag{10}$$

As we expect

$$L \cdot \mathbf{v} = \mathbf{v} + \delta\mathbf{v}, \tag{11}$$

where L is a linear transformation containing the rotation matrix, we get

$$T_1(\mathbf{v}) - T_2(\mathbf{v}) + \nabla T_2(\mathbf{v})^T \mathbf{v} = \nabla T_2(\mathbf{v})^T L\mathbf{v}, \tag{12}$$

which leads to the following linear system

$$Ax = B, \tag{13}$$

where the i^{th} row of A contains vectorized entries of $\nabla T_2(\mathbf{v_i})\mathbf{v_i}^T$, x contains the vectorized entries of L, the i^{th} entry of B is $T_1(\mathbf{v_i}) - T_2(\mathbf{v_i}) + \nabla T_2(\mathbf{v_i})^T \mathbf{v_i}$ and $\mathbf{v_i}$ is the i^{th} unit vector obtained by uniform sampling of a sphere. The embedded rotation matrix R can be recovered using the QR decomposition from L.

6.2 Surface Normal Computation

As described earlier, the surface normal, $\mathbf{n}$, at a pixel (x, y) with ABRDF T can be computed by taking a weighted geodesic mean of the normals suggested by its neighboring pixels. Let all of its P immediate neighbors be indexed $1 \dots P$ with corresponding ABRDFs as T_p, normals as $\mathbf{n}_p$ and the rotation matrices computed using the process described above as $R_1 \dots R_p$. The normal at (x, y) is then given by

$$\mathbf{n} = argmin_\mu \sum_{p=1}^{P} \frac{1}{||R_p T_p - T||^2} d^2(\mathbf{n}_p, \mu), \tag{14}$$

where $d()$ is the geodesic distance defined on the space of unit normals, arc length. We seek a geodesic mean because the domain of unit normals is the unit sphere and not the Euclidean space. This mean is also known as the weighted Karcher mean and can be computed using the following iterative scheme –

$$\mu \rightarrow exp_\mu(\epsilon\nu) \tag{15}$$

$$\nu = (1/n) \sum_{i=1}^{p} \frac{1}{||R_p T_p - T||^2} exp_\mu^{-1} \mathbf{n_p} \tag{16}$$

where exp, the exponential map, is given as

$$exp_\mu(\epsilon\nu) = \cos(|\epsilon\nu|)\mu + \sin(|\epsilon\nu|)(\nu/|\nu|) \tag{17}$$

and $exp_\mu^{-1}(n_p)$, the log map, is defined as

$$exp_\mu^{-1}(n_p) = u \cos^{-1}(\langle \mu, n_p \rangle) / \sqrt{(\langle u, u \rangle)} \tag{18}$$

where

$$u = n_p - \langle n_p, \mu \rangle \mu, \tag{19}$$

and ϵ is the iteration step size. For more details on computing means on manifolds see [52] and references therein.

6.3 Shape Recovery

Once the normal field has been computed, we use one of the standard techniques ([48]) to recover the surface. If $z(x, y)$ defines the surface, the normal at a location (x, y) is given by $(z_x \; z_y \; -1)^T$ where z_x and z_y denote the partial derivatives of the surface with respect to x and y. If $(n_x \; n_y \; n_z)^T$ denotes the surface normal at location (x, y), we have the following relations

$$z_x = -n_x / n_z \tag{20}$$

$$z_y = -n_y / n_z. \tag{21}$$

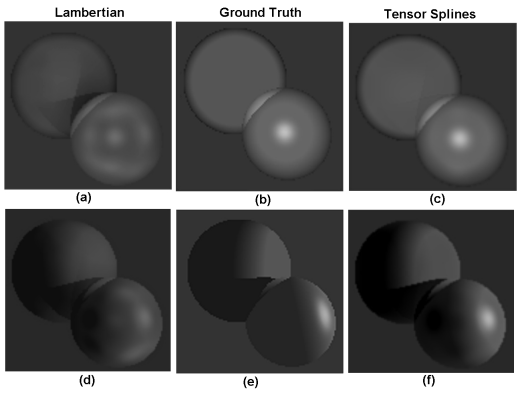

Fig. 4. **Synthetic Example**. Cast shadows and specularities are notedly better rendered using Tensor Splines than the Lambertian Model.

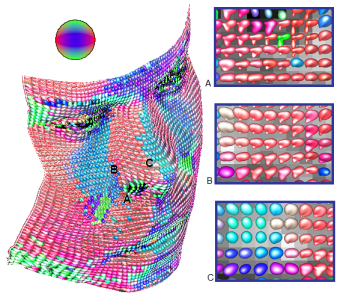

Fig. 5. Recovered ABRDFs for a human face. Complex shapes of ABRDFs in various regions of the face can be readily noted. Functions are colored according on their maximal value direction. The color to direction mapping is shown on top left.

Using the forward difference approximation of the partial derivatives we obtain the following two equations

$$n_z z(x+1, y) - n_z z(x, y) = n_x \tag{22}$$

$$n_z z(x, y+1) - n_z z(x, y) = n_y, \tag{23}$$

which provide a linear relation between the surface values at the grid point and the known surface normals. The surface can now be recovered by solving an over-determined system of linear equations. At the boundary points, $n_z \approx 0$ and hence the above formulation is not valid. A single linear constraint is instead obtained by eliminating n_z in Eq. 22 and 23 above

$$n_x z(x, y) - n_x z(x, y+1) = n_y z(x+1, y) - n_y z(x, y). \tag{24}$$

6.4 Novel Pose Relighting

With the facial shape in hand, novel poses can be rendered by simply changing the viewpoint. But generating novel illumination conditions in the novel pose is not trivial as the ABRDFs estimated from a different pose cannot be directly used. If the ABRDF field was estimated in pose P_1 and if we wish to generate an image with a novel illumination in a new pose P_2, we have

Fig. 6. Images synthesized using Tensor Splines under novel illumination direction (mentioned on each image as (azimuth,elevation)). 9 images used as input were illuminated from (-20,60),(0,45),(20,60),(-50,0),(0,0),(50,0),(-50,-40),(0,-35) and (50,40) directions.

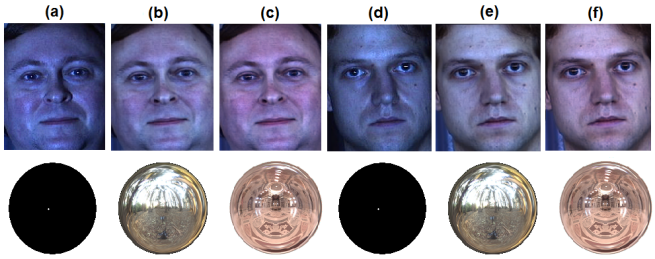

Fig. 7. Images relit with complex lighting. 1^{st} image of both subjects is lit by a point source while the next two are lit by Eucalyptus Grove and St. Peter's Basilica light probes respectively. Light probes are provided below the facial images.

to rotate the ABRDFs by the same rotation which is required to change P_1 to P_2. Once the orientations of the BRDFs have been rectified, images of the face in the new pose with novel illumination can be generated by evaluating the ABRDF field in the desired directions.

We would like to point out that the specularities are view dependent and accurately speaking, cannot be directly transferred from one pose to another. Most of the existing Lambertian methods ignore this effect but the few who deal with this problem, handle it by either explicitly obtaining the specular component by using polarized lighting (e.g. [14], [23]), which required specialized data acquisition, or by assuming a parametric form for the specular component of lighting (e.g. [20]).

Our Cartesian tensor representation for ABRDF does not discriminate against specularities and estimates the ABRDF as best as possible from the available intensity values. Thus it should be possible to recover and manipulate the specular component separately, but at this stage, we have made the assumption that specularities

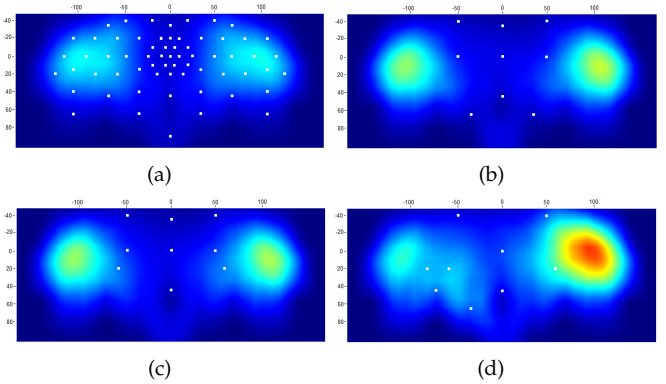

Fig. 8. Distribution of errors as the configuration of input changes. X-axis represents azimuth and Y-axis represents elevation angles. Hotter colors show larger errors. The white dots represent the exact directions of illumination in images used as input.

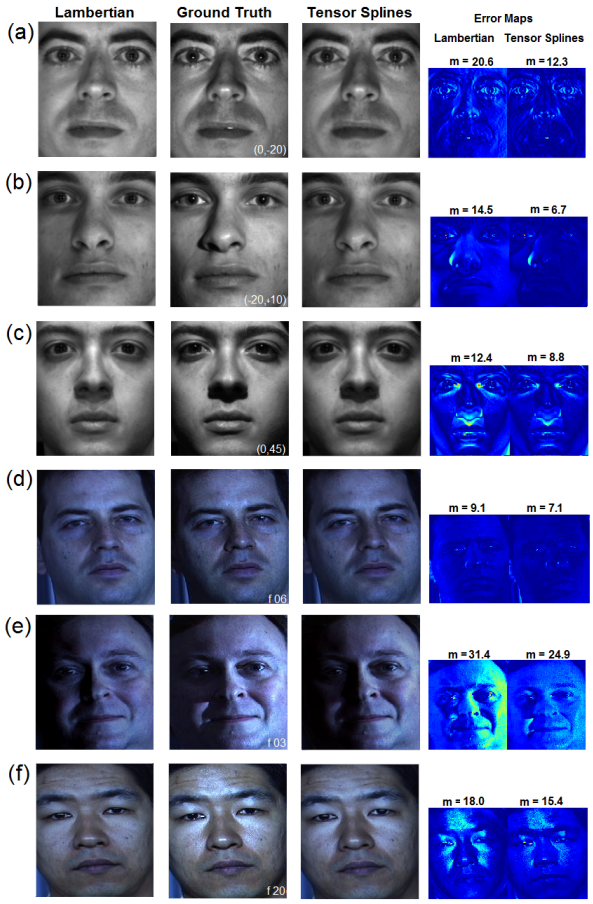

Fig. 9. **Relighting Comparison**. Angle of illumination is shown in the lower right corner of the ground truth images. '**m**' is the mean intensity error per pixel. Hotter colors in the error maps indicate larger errors.

do not change drastically across facial poses. The validity of this assumption is supported by the results presented in the next section.

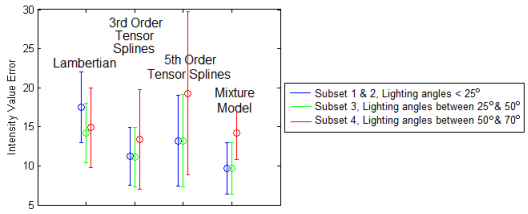

Fig. 10. Per pixel intensity error comparison.

7 Experimental Results

In order to evaluate the proposed method for relighting and shape recovery, we conducted several detailed experiments which are presented here. Since it has been shown for the popular Lambertian model that the space of images with illumination variation can be approximated quite accurately using a 9 dimensional subspace ([12], [13]), we have taken on the challenge of also working with just 9 image. Note that with 9 samples of the ABRDF field and the splines based smoothness constraint, at most 10 coefficients can be recovered and hence our central results use bi-cubic 3^{rd} order Tensor Splines. On a Dual Core AMD Opteron 2.2 GHz machine with 4GB memory it takes around 10 second to complete 40 iterations, which provide reasonably good results.

The experiments were carried out on the Extended Yale B [40] (28 subjects, in 9 poses and 64 illumination conditions) and the CMU PIE [49] (68 subjects in 13 poses and 43 illumination conditions) benchmark databases. Note that CMU PIE has 21 usable point source illuminated images while in Extended Yale B all 64 illuminations are point source. Qualitatively, we would like to point out that the CMU PIE dataset has a very narrow band of illumination variation above and below the equator and Extended Yale B has inconsistencies across pose change.

7.1 Relighting

Synthetic Example: Foremost, in Fig. 4 we demonstrate the expressibility of the Tensor Splines model over the Lambertian Model using an artificially rendered scene. The larger sphere is Lambertian while the smaller sphere is more specular. The two cases (b) and (e) are chosen to demonstrate the impact of cast shadows and specularities on both the proposed and the Lambertian models. Nine images in each case were used to estimate the intensity variation at each pixel using the two models and then images under novel illumination directions were rendered in cases (a), (c), (d) and (f). It can be noted that the Tensor Splines model captures specularities and cast shadows to a larger extent than the Lambertian Model when the same amount of input data is used.

ABRDF shape: Next we demonstrate that Tensor Splines can capture non-trivial shapes of facial ABRDFs. In Fig. 5 we show the ABRDF field of a subject from the Extended Yale B database estimated using 9 images.

Fig. 11. Detailed pose variation with texture-less in upper right and depth-map in lower right. Photo-realistic renderings even in extreme poses can be noticed.

Three different regions of the face have been shown in detail where complicated shapes of the ABRDF can be noticed. Regions A and B have more complicated ABRDFs because these have to accommodate shadows. The spherical functions in the image have been color coded based on their maximal value direction. Mapping of directions to colors is provided on the top left of the figure.

Novel Relighting: In Fig. 6 four different subjects lit in various novel point source illuminations are depicted. For the first two rows the illumination direction varies across the azimuth angle while in the next two rows the variation is in the elevation angle. It can be noticed that our method can accurately interpolate as well as extrapolate from the images provided as input. Further, difficult effects like cast shadows and specularities have been photo-realistically rendered without using any additional ray tracing.

Complex Lighting: Since our technique can estimate the entire ABRDF field, it can be easily used to render images in complex lighting conditions. In Fig. 7 we present such results for two subjects from the CMU PIE database. Below each face is its lighting condition. Images (a) and (d) show one of the nine images used in ABRDF estimation. The next two images for each of the subjects are lit by light probes ([50]) named Eucalyptus Grove and St. Peter's Basilica respectively. For color images we estimate the ABRDF field for each channel separately. The images were relit by taking a weighted combination of the point source lit images. We used 2500 samples of the light probe to render these images.

Impact of Input Image Distribution: To examine the impact of the distribution of input images when the order of tensor is fixed (3^{rd} order in this case) we use the Extended Yale B dataset in Fig. 8. To set a baseline we estimated the ABRDF field for 10 subjects using all the 64 images as input, rendered images in the same 64 direction and computed the total error with respect to the ground truth in part (a). Errors were similarly computed for three other cases with only 9 input images, but in different configurations. In cases (b) and (c) input image illumination was uniformly distributed in front of the face while in case (d) images with the direction

biased toward one side were used. To visualize the results, we color coded larger errors with hotter colors and plotted them as a continuous images. It can be noted that when all 64 images are used as input, case (a), the error is the least. For the 9 image cases (b) and(c), where the illumination directions in the input images are uniformly distributed, the error is more than case (a) but notedly less than the case when distribution is skewed in one direction, case (d). Hence, as expected, our method performs better when the input images are uniformly sampled from the sphere. Moreover, the errors in all cases are concentrated towards the extreme illumination angles and for near frontal illumination condition the performance is not particularly affected by the input image distribution.

Comparison with the Lambertian Model: Next we present both qualitative and quantitative comparison between the proposed model and the Lambertian model in Fig. 9. Parts (a)-(c) present results on Extended Yale B while parts (d)-(f) present results on CMU PIE database. Next to the Lambertian, Ground Truth and Tensor Splines results, error maps depicting the pixel-wise errors in the Lambertian and Tensor Splines cases are shown. It can be readily noted that the Tensor Splines Model is successful in capturing the true appearance to a larger extent than the Lambertian Model.

Comparison with the Validation Model: We present a quantitative comparison among the variants of our method, the Lambertian model and the validation model (Section 5) in Fig. 10. A natural question that arises is why should an order 3 Cartesian tensor be suitable for estimating the ABRDFs? To answer this question, we computed the average intensity error per pixel over all 38 subjects in 64 illumination directions of the Extended Yale B dataset using the Lambertian model, 3^{rd} order Tensor Splines, 5^{th} order Tensor Spline and the mixture of single lobed functions (Eq. 6). All 64 illumination directions were used for the mixture models (on account of large number of unknowns) while for the other three only 9 images were used. For the mixture model, we chose μ_i values using a dense uniform sampling (642 directions) of the unit sphere obtained by the 4^{th}-order tessellation of an icosahedron. We have presented results shattered along the standard subsets of the Extended Yale B database. It can be noted that the error for subset with extreme lighting (subset 4) is more than that of other sets, for all methods. More importantly, even with a considerably large amount of data and a very flexible estimation model, the errors obtained from the mixture model is quite similar to those obtained from the 3^{rd} order Tensor Splines. This indicates that though a 3^{rd} order Tensor Spline can only accommodate three lobes, for most facial ABRDFs this suffices. The 3^{rd} order Tensor Spline outperforms the Lambertian model and even the 5^{th} order Tensor Spline, which suggests possible over-fitting in the 5^{th} order model.

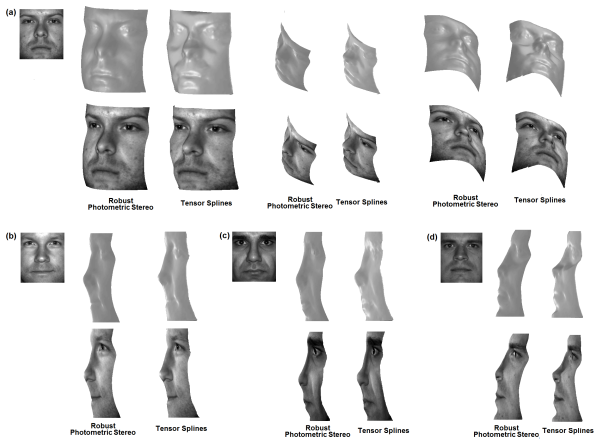

Fig. 12. Shape comparison with the Robust Photometric Stereo on Extended Yale B dataset.

7.2 Estimating Shape

Before presenting results we would like to make an important observation that the proposed method recovers normal at each pixel by taking local neighborhood into account. This is at contrast with both the Lambertian Photometric Stereo ([54]), which works at each pixel independently and Morphable Models [9], which impose a global prior on shape. Our method can be looked at as being somewhere in between these two extremes.

Range of Pose Change: Since the primary objective of our shape recovery method is to render images of faces in novel poses, we present various images of a face in a range of different pose in Fig. 11. The lighting is held fixed as the poses changes from right profile and left profile and from above the plane view to below that plane view. The ABRDF field for this subject was recovered using 9 images under the illumination configuration shown in Fig. 8(b). The recovered shape without any texture is presented on the right with the depth map below it. It can be noted that the rendered images are very photo-realistic, even in extreme poses. Since we capture the ABRDF only for the frontal hemisphere, while relighting using Tensor Splines model, the maximum angle between the viewing direction (pose) and the illumination direction can be 90°.

Comparison with Robust Photometric Stereo: We present face shapes estimated by our method and the Robust Photometric Stereo [54] (9 input images for both methods) in Fig. 12. Results for four different subjects, both with and without texture, are presented. Based on the results following conclusions can be drawn: First, since the Tensor Splines method imposes local smoothness, the recovered shape lacks some minute details like the mole on the chin in case (a) as compared to the Robust Photometric Stereo. Second, since the Tensor Spline more seamlessly handles cast shadows and specularities as compared to Robust Photometric Stereo, regions affected by cast shadows and specularities, specially the nose, it is better recovered by Tensor Splines. This can be readily noted in cases (a), (c) and

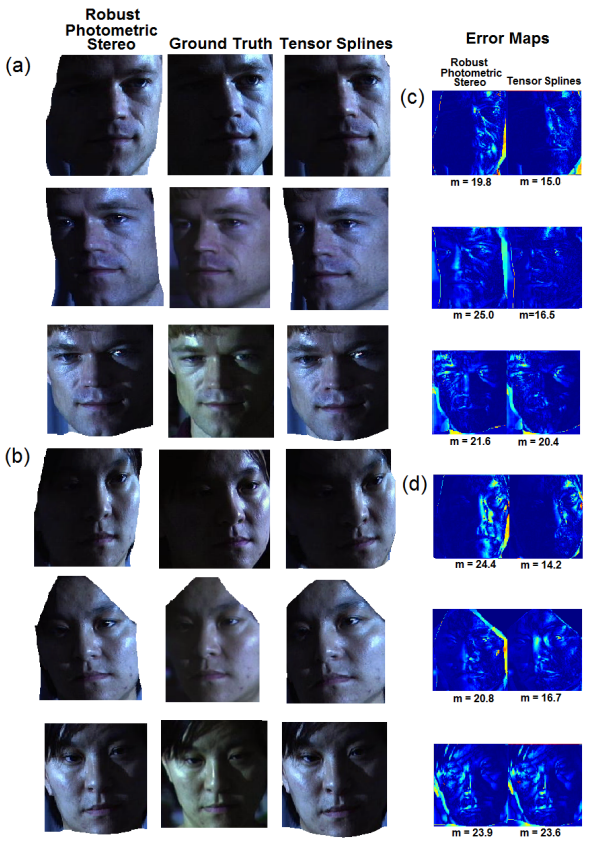

Fig. 13. Shape comparison with the Robust Photometric Stereo and the ground truth on CMU PIE dataset. '**m**' is the mean intensity error per pixel. Hotter colors in the error maps indicate larger errors. The azimuth angle for the half vectors in input were $15°$, $15°$ and $-17°$ for three cases respectively for both the subjects. The same angles for novel images in case (a) were $43°$, $16°$ and $-27°$ and in case (b) $43°$, $22°$ and $-31°$.

(d). Finally, the Tensor Splines method seems to have better global shape estimation. For instance, in case (a), the shape recovered by Robust Photometric Stereo is titling backwards towards the top. In case (c), the region around the mouth seems unnaturally warped in Robust Photometric Stereo while in case (d), relative positioning of nose and eyes seems more realistic using Tensor Splines. In summary, these results demonstrate that Tensor Splines may lack minute details but models facial features in a more photo-realistic fashion than Robust Photometric Stereo, when the input images have cast shadows and specularities. When the illumination direction in the input images makes small angle with the viewing direction i.e. when there is no significant shadowing, Robust Photometric Stereo can be useful.

Comparison with Ground Truth: We used simultaneously obtained multiple poses from CMU PIE database to compare the pose changed images generated using Tensor Splines and the Robust Photometric Stereo with the ground truth. Results are presented in Fig. 13 which are used to demonstrate both the quality of shape recovery as well as the impact moving specularities from one

Fig. 14. Simultaneous pose and illumination variation.

pose to another. In cases (a) and (b) intensity images are presented where visually the results can be compared. In cases (c) and (d), pixel wise error maps for cases (a) and (b) respectively are presented. It can be noted that the errors in global shape estimation in Robust Photometric Stereo hampers it significantly as compared to Tensor Splines. It is also evident that moving specularities from one pose to another does not create visually significant artifacts, since the texture images for both Tensor Splines and Robust Photometric Stereo were taken in a different pose than the ones shown in (a) and (b). For all cases, the shape was recovered using the luminance channel.

Simultaneous Pose and Illumination Variation: In Fig. 14 one subject each from CMU PIE and Extended Yale B database are shown in various novel poses and novel illumination conditions. The ABRDF fields for both cases were recovered using 9 frontal pose images, and the shape for the color images were recovered using the luminance channel. With the change of pose we have retained the ABRDF field learnt using the front pose but it can noted that the results are photo-realistic even when specularities are not explicitly modified and transferred.

7.3 Face Recognition

Face recognition is one of most popular applications of facial image analysis. This problem is usually defined as finding the closest match of a given probe (facial) image from a set of gallery (facial) images. The complexity of the recognition problem increases as the gallery set gets less and less representative of the expected probe images. Thus, augmenting the gallery set with meaningful images can aid recognition.

In the case of illumination invariant face recognition, if the gallery set has nine or more images with known illumination directions, the proposed method can be used to generate realistic images with a much denser sampling of the illumination space. This database augmentation can potentially improve performance of any classifier. Here we have demonstrated the usefulness of the proposed scheme using the very simple Nearest-Neighbor classifier.

We have used Extended Yale B data for this experiment. The database is divided into 4 subsets with the lighting getting more and more extreme as we go from subset 1 to 4. The difficulty of classifying images from these subsets also increases in a similar fashion. The obtained recognition error rate are reported in Table

2 where we have also presented results reported by existing methods and their respective references are listed. Results for the first seven techniques were taken from [7] and the rest were taken from the respective references. Along with the error rates, we have also listed the number of images required by each method in the gallery set. For our method we used the nine images in the configuration shown in Fig. 8(b). It can be noted that even with the naive nearest neighbor classification strategy our method produces near perfect results.

TABLE 2
Face recognition errors rates. N: # of input images.

Method	N	Subset 1&2	Subset 3	Subset 4	Total
Correlation [2]	4	0.0	23.3	73.6	29.1
Eigenfaces [8]	6	0.0	25.8	75.7	30.4
Linear subspace [1]	7	0.0	0.0	15.0	4.7
Cones-attached [10]	7	0.0	0.0	8.6	2.7
Cones-cast [10]	7	0.0	0.0	0.0	0.0
9PL [40]	9	0.0	0.0	2.8	0.8
3D SH [7]	1	0.0	0.0	2.8	0.8
Harmonic (SFS) [6]	1	0.0	0.0	12.8	4.0
Tensor Splines	9	0.0	0.0	1.6	0.5

8 Conclusions, Limitations and Future Work

In this paper we have presented a novel comprehensive system for capturing the reflectance properties and shape of the human faces using Tensor Splines. Since our method requires at least 9 input images with known illumination directions, we fall short of the ideal solution described in the introduction, but show an improvement over the popular Lambertian model. Accurate recovery of ABRDF field from a single image with cast shadows and specularities with no lighting information remains a challenge. The central problem in the single image case stems from the dearth of information to constrain the space of all possible ABRDF fields. Use of strong prior information presents itself as a potentially effective way to constrain the search space, but attempts so far (e.g. [9]) suffer from the need of manual intervention and cumbersome computational requirements. We would like to explore the use of Tensor Spline ABRDF fields as prior information to meaningfully predict ABRDF fields using single input images in future. Use of a shape prior can also potentially aid in shape recovery.

While relighting images in novel poses, we make the assumption that the ABRDF field maintains the same specular information across poses. Though practically useful, this is not fully valid. We have dealt with the specularities in a data driven fashion but possible attempt can be made to explicitly model the specularities, which we would like to explore in future. It should be noted that though the problem of detecting specularities is relatively well studied, the problem of realistically predicting specularities in novel poses without using specialized imaging tricks (like special filters) remains challenging. Possible improvement can also be made in our model by incorporating non-uniform smoothness as opposed to the current setup. In our relighting framework, the angle between the viewing direction (pose) and the illumination direction must be less than 90°. Note that our method aims to generate individual images that are photorealistic and does not attempt to tackle the more complicated problem of animation, where temporal component is also involved.

Besides the relighting and pose change applications described in the paper, our technique can also be used for image up-sampling and compression. The former is possible because the Tensor Splines representation creates a continuous field of ABRDF coefficients across the image, which can be sampled at a sub-pixel resolution. The later exploits the capability of ABRDFs to represent images of a face under infinitely many lighting directions using just a few coefficients.

In conclusion, the Tensor Splines framework for the analysis and modeling of illumination and pose variation of facial images provides a useful alternative to the Lambertian assumption. It also seems that the collective analysis of shape and reflectance through ABRDFs is promising as an alternative to separate facial BRDF and shape analysis.

Acknowledgment

A part of the work presented here on relighting appeared in the Proceedings of IEEE CVPR 2008 [26]. This research was in part funded by the University of Florida Alumni Fellowships to Ritwik Kumar and Angelos Barmpoutis.

References

[1] A. U. Batur and M. Hayes, "Linear subspaces for illumination robust face recognition", CVPR, 2:296-301, 2001.
[2] T. Brunelli, R. Poggio, "Face recognition: features versus templates", PAMI, 15(10):1042-1052, Oct 1993.
[3] Z. Wen, Z. Liu, T. Huang,"Face Relighting with Radiance Environment Maps", CVPR, page 158–165, 2003.
[4] M. Chandraker, S. Agarwal, D. Kriegman, "ShadowCuts: Photometric stereo with shadows", CVPR, 2007.
[5] S. K. Zhou, G. Aggarwal, R. Chellappa, D. W. Jacobs, "Appearance Characterization of Linear Lambertian Objects, Generalized Photometric Stereo, and Illumination-Invariant Face Recognition", PAMI, 2007.
[6] W. Smith, E. R. Hancock, "Facial Shape-from-shading & Recognition Using Principal Geodesic Analysis & Robust Statistics", IJCV, 2008.
[7] L. Zhang and D. Samaras, "Face Recognition from A Single Training Image under Arbitrary Unknown Lighting using Spherical Harmonics", PAMI, 28(3):351–363, March 2006.
[8] P. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions", CVPR, pp. 995-999, 1994.
[9] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces", SIGGRAPH, pp. 187-194, 1999.
[10] A. Georghiades, P. Belhumeur and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition Under Variable Lighting and Pose", PAMI, 23(6):643–60, 2001.
[11] R. Basri, D. Jacobs and I. Kemelmacher,"Photometric Stereo with General, Unknown Lighting", IJCV, 72(3): 239–257, 2007.
[12] R. Basri and D. Jacobs, "Lambertian reflectances and linear subspaces", PAMI, 25(2):218–233, 2003.
[13] R. Ramamoorthi and P. Hanrahan, "A Signal-Processing Framework for Inverse Rendering", SIGGRAPH, pp. 117–128, 2001.
[14] T. Zickler, R. Ramamoorthi, S. Enrique, P. N. Belhumeur, "Reflectance Sharing: Predicting Appearance from a Sparse Set of Images of a Known Shape", PAMI, 2006.
[15] N. Alldrin, D. Kriegman, "Shape from Varying Illumination and Viewpoint", ICCV, 2007.
[16] Neil Alldrin, Todd Zickler, D. Kriegman, "Photometric Stereo With Non-Parametric and Spatially-Varying Reflectance", CVPR, 2008.

[17] A. Hertzmann, S. M. Seitz, "Shape and materials by example: A photometric stereo approach", CVPR, 2003.
[18] A. Hertzmann, S. M. Seitz, "Example-based photometric stereo: Shape reconstruction with general, varying brdfs", PAMI, 2005.
[19] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, "Illumination-Based Image Synthesis: Creating Novel Images of Human Faces Under Differing Pose & Lighting", IEEE Workshop on Multi-View Modeling & Analysis of Visual Scenes, 1999.
[20] A. S. Georghiages, "Recovering 3-D Shape and Reflectance From a Small Number of Photographs", Eurographics, 2003.
[21] D. Goldman, B. Curless, A. Hertzmann, S. Seitz, "Shape & Spatially-Varying BRDFs from Photometric Stereo", ICCV, 2005.
[22] J. Lee, B. Moghaddam, H. Pfister and R. Machiraju, "A Bilinear Illumination Model for Robust Face Recognition", ICCV, 2005.
[23] P. Debevec, T. Hawkins, C. Tchou, H. Duiker, W. Sarokin, M. Sagar "Acquiring the Reflectance Field of a Human Face", SIGGRAPH, 2000.
[24] S. Magda, T. Zickler, D. J. Kriegman, P. N. Belhumeur, "Beyond Lambert: Reconstructing Surfaces with Arbitrary BRDFs", ICCV, 2001.
[25] W. Zhao, R. Chellappa,"Symmetric Shape from Shading Using Self-Ratio Image", IJCV, Vol. 45, pp. 55–75, 2001.
[26] A. Barmpoutis, R. Kumar, B. C. Vemuri, A. Banerjee,"Beyond the Lambertian Assumption: A generative model for Apparent BRDF fields of Faces using Anti-Symmetric Tensor Splines", CVPR, 2008.
[27] S. Biswas, G. Aggarwal, R. Chellappa, "Robust Estimation of Albedo for Illumination-invariant Matching & Shape Recovery", ICCV, 2007.
[28] K. C. Lee and B. Moghaddam, "A practical face relighting method for directional lighting normalization", IWAMFG, 2005.
[29] A. Shashua, T. R.-Raviv, "The quotient image: Class-based re-rendering & recognition with varying illumination", PAMI, 2001.
[30] R. Gross, I. Matthews, S. Baker, "Appearance-Based Face Recognition and Light-Fields.", PAMI, 26(4):449–465, 2004.
[31] H. Chan, W. W. Bledsoe, "A Man-Machine Facial Recognition System-Some Preliminary Results", Panoramic Research Inc, 1965.
[32] A Shashua,"On Photometric Issues in 3D Visual Recognition from a Single 2D Image" IJCV, vol. 21, pp. 99–122, 1997.
[33] A.L. Yuille, D. Snow, R. Epstein, P. N. Belhumeur, "Determining Generative Models of Objects under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability", IJCV, 1999.
[34] P.N. Belhumeur, D.J. Kriegman, "What Is the Set of Images of an Object under All Possible Illumination Conditions?" IJCV, 1998.
[35] T. Malzbender, D. Gelb, H.Wolters. "Polynomial Texture Maps", SIGGRAPH, pages 519-528, 2001.
[36] R. Ramamoothi, P. Hanrahan, "On the Relationship between Radiance and Irradiance: Determining the Illumination from Images of a Convex Lambertian Object," JOSA, vol. 18, 2001.
[37] P. Belhumeur, D. Kriegman, A. Yuille, "The Bas-Relief Ambiguity", IJCV, 35(1), 1999, pp. 33–44
[38] J. P. O'Shea, M. S. Banks, M. Agrawala, "The Assumed Light Direction for Perceiving Shape from Shading", ACM Applied Perception in Graphics and Visualization, 2008.
[39] Z. Liu, Y. Shan, and Z. Zhang, "Expressive expression mapping with ratio images", SIGGRAPH, pages 271-276, 2001.
[40] K. Lee, J. Ho, D. Kriegman, "Acquiring Linear Subspaces for Face Recognition under Variable Lighting", PAMI, 27, 2005.
[41] G. Borshukov, J.P. Lewis, "Realistic Human Face Rendering for "The Matrix Reloaded", Sketch, SIGGRAPH, 2003.
[42] S. C. Foo. "A gonioreflectometer for measuring the bidirectional reflectance of material for use in illumination computation", Masters thesis, Cornell University, Ithaca, 1997.
[43] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, M. Gross, "Analysis of Human Faces using a Measurement-Based Skin Reflectance Model", SIGGRAPH, 2006.
[44] C. de Boor, "On Calculating with B-Splines", Journal of Approximation Theory, 6:50-62, 1972.
[45] C. Lawson, R. J. Hanson, "Solving Least squares problems", Prentice-Hall, Englewood Cliffs, 1974.
[46] B.T. Phong, "Illumination for computer generated images", Communications of the ACM, 18(6):311-317, June 1975.
[47] K.E. Torrance, E.M. Sparrow, "Theory for off-specular reflection from roughened surfaces", J. Opt. Soc. Am., 57, 1967.
[48] B.K.P. Horn, "Robot Vision", MIT Press, 1986.
[49] T. Sim, S. Baker, M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces", Technical Report CMU-RI-TR-01-02, Robotics Institute, CMU, 2001.
[50] P. Debevec, "Rendering Synthetic Objects Into Real Scenes: Bridging Traditional & Image-Based Graphics With Global Illumination & High Dynamic Range Photography", SIGGRAPH 98.
[51] T. P. Wu, K. L. Tang, C. K. Tang, T. T. Wong, "Dense Photometric Stereo: A Markov Random Field Approach", PAMI, 28(11), 2006.
[52] A. Barmpoutis, B. C. Vemuri, T. M. Shepherd, J. R. Forder, "Tensor splines for interpolation and approximation of DT-MRI with applications to segmentation of isolated rat Hippocampi", IEEE Transactions on Medical Imaging, 2007.
[53] M. L. Koudelka, P. N. Belhumeur, S. Magdha, D. J. Kriegman, "Image-based Modeling and Rendering of Surfaces with Arbitrary BRDFs", IEEE CVPR, 2001.
[54] E. N. Coleman, Jr, R. Jain, "Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry", Computer Graphics and Image Processing, 1982.
[55] I. Kemelmacher, R. Basri, "Molding Face Shapes by Example", ECCV, 2006.
[56] H. Jeffreys, "Cartesian Tensors", Cambridge:The Univ. Press, 1931.
[57] K. Yoon, E. Prados, P. Sturm, "Joint Estimation of Shape and Reflectance using Multiple Images with Known Illumination Conditions", IJCV, 2009.

Ritwik Kumar received his B.Tech. degree in Information and Communication Technology from Dhirubhai Ambani Institute of Information and Communication Technology (DAIICT), India in 2005. Since 2005 he has been a Ph.D. student at the Center for Vision, Graphics and Medical Imaging at the Dept. of Computer and Information Science and Engineering at the University of Florida, Gainesville, FL, USA. His research interests include machine learning, computer vision and medical image analysis. He is a recipient of DAIICT President's Gold Medal (2005) and the University of Florida Alumni Fellowship (2005 - 2009).

Angelos Barmpoutis received the B.Sc. in Computer Science from the Aristotle University of Thessaloniki in 2003, the M.Sc. in Electronics and Electrical Engineering from the University of Glasgow in 2004, and the Ph.D. degree in Computer and Information Science and Engineering from the University of Florida in 2009. He has coauthored more than 20 journal and conference publications. His current research interests lie in the areas of biomedical image processing, machine learning and computer vision.

Arunava Banerjee is an Associate Professor in the Department of Computer and Information Science and Engineering at the University of Florida, Gainesville. He received his M.S. and Ph.D. degrees in the Department of Computer Science at Rutgers, the State University of New Jersey in 1995 and 2001, respectively. His research interests include algorithms, computer vision, operational research, and computational neuroscience. He received the Best Paper award at the Nineteenth International Conference on Pattern Recognition (ICPR), 2008.

Baba C. Vemuri received the PhD degree in electrical and computer engineering from the University of Texas at Austin in 1987. He joined the Department of Computer and Information Sciences at the University of Florida, Gainesville, in 1987 and is currently a Professor and Director of the Center for Vision, Graphics & Medical Imaging. He was a coprogram chair of the 11th IEEE International Conference on Computer Vision (ICCV 2007). He has been an AC and a PC member of several IEEE conferences. He was an AE of IEEE TPAMI (from 1992-96) and the IEEE TMI (from 1997 -03). He is currently an AE for the Journal of MedIA and CVIU. He is a Fellow of the ACM and the IEEE. His research interests include medical image analysis, computational vision, and information geometry.