REP-2009-467: Non-Lambertian Reflectance Modeling and Shape Recovery for Faces using Anti-Symmetric Tensor Splines

Ritwik Kumar, *Student Member, IEEE,* Angelos Barmpoutis, *Student 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 anti-symmetric tensor spline based framework for facial image analysis. We show that using this framework, 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 anti-symmetric tensor spline model using a more general continuous 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—Anti-Symmetric Tensor Splines, Non-Lambertian Reflectance, 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 accurate modeling and realistic 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 an epitome 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 facial surface has come a long way since we started [34], with many impressive strides along the way (e.g. [12], [25], [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 required any manual intervention and should not be fazed by 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 existing state-of-the-art is largely comprised of systems which relaxes one or more of these conditions while satisfying others. Common simplifying assumptions include applicability of 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 amount of data obtained from custom built rigs (e.g. [25]) 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 real 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 spatially varying non-Lambertian

This manuscript is under review at IEEE Transactions on Pattern Analysis and Machine Intelligence
All authors are with the Department of Computer and Information Science and Engineering at the University of Florida, Gainesville, FL 32611, USA

reflectance model. Unlike most techniques, our method does not require input images to be free of cast shadows or specularities and can reproduced these in the novel renderings. It does not require any manual initialization and is a purely image based technique (no expensive 3D 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 [1] is based on a novel framework of anti-symmetric tensor splines which can be used to approximate any n-dimensional field of spherical functions. In the case of faces, we use anti-symmetric tensor splines to approximate the field of Apparent Bidirectional Reflectance Distribution function (ABRDF) for a fixed viewing direction. Unlike the BRDF, the ABRDF at each pixel captures the variation in intensity as a function of illumination and viewing direction and thus is 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 computing the weighted Karcher mean of the rotated normals of the neighborhood pixels. To recover the requisite rotation we present an extremely efficient linear method along with the more straightforward non-linear optimization based method. Once the rotations for the ABRDF field have been determined, we initialize the normals with the maxima of the ABRDF functions and iteratively compute the surface normals. 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 pose with novel illumination conditions can then be rendered while seamlessly accounting for attached as well as cast shadows.

The rest of the paper is organized as follows: In Section 2 we present a detailed survey of related work. In Section 3, we give an overview of our system and in Section 4, the novel Anti-Symmetric Tensor Spline framework for ABRDF modeling is introduced. In Section 5, we present a validation model for the Anti-Symmetric Tensor Spline method. In Section 6, the novel shape recovery procedure is presented and in Section 7, we present detailed experimental results and related discussion. Finally we conclude in Section 8.

2 Related Work

The sheer size of the facial shape-reflectance modeling literature allows its taxonomy to be carried along various possible lines. Here we have classified methods based on various assumptions made by them while modeling facial reflectance and shape. We have also summarized few of the key methods along with associated assumption in Table 1.

1. A part of the work presented here on relighting appeared in the Proceedings of IEEE CVPR 2008 [29].

A large fraction of the existing techniques for facial image analysis work with the Lambertian assumption for reflectance. This translates to assuming that the BRDF at each point on the object's surface has the same shape, that of half cosine function, which has been scaled by a constant – albedo, and is oriented about 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 [35]. When ambient lighting component is included, this subspace expands to become 4-dimensional [36] and when attached shadows are taken into account, the subspace grows to become infinite dimensional – illumination cone [37].

Spherical harmonic analysis of the Lambertian kernel has shown that even though the illumination cone is infinite dimensional, it can be approximated quiet well by a lower dimensional subspaces ([39], [13], [12]). In particular, these methods can produce impressive results with 9 basis images, though they require 3D shape and albedo field as input. These basis images can also be directly acquired using the "universal virtual" lighting conditions [43]. 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 feature to initialize a complicated non-linear optimization which can take long time to converge and can suffer from local minima. Using the idea of low dimensional subspace explored above, [33] 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, majority of the face databases still remain 2D and hence it is more pragmatic to work with 2D images alone [24]. Methods that are purely image based and work with the Lambertian assumption generally apply photometric 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 [21] 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 [28] and [41]. As these methods work with just a single image, besides requiring absence of cast shadows in input, they make additional assumptions

TABLE 1
Requirements, Assumptions and Capabilities of Existing Methods

Methods	Assumed Surface BRDF Model	No. of Images as Input	Relit Images Presented	Shape or Pose Results Presented	Cast Shadow Allowed in Input	Purely Image based (No 3D Scans)	Other Assumptions, Requirements and Limitations
1999 MVIEW Georgihades et al.[21]	Lambertian	≥ 3	✔	✔	✘	✔	Near frontal illumination expected, Ray tracing for cast shadows.
1999 SIGGRAPH Blanz et al.[9]	Phong	1	✔	✔	✘	✘	No attached shadows, Manual initialization to fit 3D model.
2000 SIGGRAPH Debevec et al.[25]	Non-Lambertian	≥ 2000	✔	✔	✔	✔	Custom rig for data collection, Structured lighting for shape.
2001 SIGGRAPH Ramamoorthi et al.[13]	Lambertian	≥ 3	✔	✘	✘	✘	Distant and isotropic lighting, 3D Scans needed as input.
2001 SIGGRAPH Malzbender et al.[38]	Non-Lambertian	≥ 50	✔	✔	✘	✘	Custom rig for data acquisition, No specularity allowed in input.
2001 PAMI Georgihades et al.[10]	Lambertian	≥ 7	✔	✔	✘	✔	Almost no attached shadow, Symmetric faces, Ray tracing.
2001 PAMI Shashua et al.[10]	Lambertian	1	✔	✘	✘	✔	Bootstrap set of images required, Ideal class assumption.
2001 IJCV Zhao et al.[28]	Lambertian	1	✔	✔	✘	✔	No attached shadows, Symmetric faces, piecewise constant albedo.
2001 ICCV Magda et al.[26]	Non-Lambertian	≥ 300	✘	✔	✔	✔	Known lighting directions, Lighting should doubly cover the directions
2003 EGSR Georghiades et al.[22]	Non-Lambertian	≥ 12	✔	✔	✘	✔	3 sources/pixel, ad-hoc shadow detection, Spatially constant BRDF.
2003 CVPR Wen et al.[4]	Diffuse	1	✔	✔	✘	✘	Symmetric lighting, 3D Model Fitting, Manual initialization.
2003 PAMI Basri et al. [12]	Lambertian	1	✔	✘	✘	✘	Distant & isotropic lighting, 3D scans required, Manual initialization.
2004 PAMI Gross et al. [33]	Lambertian	≥ 1	✘	✔	✘	✔	Manual delineation of feature points for better recognition.
2005 ICCV Goldman et al. [23]	Non-Lambertian	12	✔	✔	✘	✔	Known lighting, HDR images expected, Manual threshold selection
2005 ICCV Lee et al. [24]	Non-Lambertian	1	✔	✔	✔	✘	Custom data acquisition, 3D Model fitting with manual initialization.
2005 PAMI Hertzmann et al. [20]	Non-Lambertian	≥ 8	✘	✔	✘	✔	No shadows expected, Reference object expected, Symmetry of faces.
2005 IAMF Lee et al. [31]	Lambertian	1	✔	✘	✔	✘	Shadowed pixel gets default albedo, Universal 3D face model required.
2006 PAMI Zhang et al. [7]	Lambertian	1	✔	✔	✘	✘	3D Model Fitting with manual initialization.
2006 PAMI Zickler et al. [14]	Non-Lambertian	≥ 1	✔	✘	✘	✘	Point lighting sources with known directions, Object shape required.
2007 CVPR Chandraker et al. [5]	Lambertian	≥ 4	✘	✔	✔	✔	3 sources/pixel, Known lighting, Normals can't be on bisection planes.
2007 ICCV Alldrin et al. [16]	Non-Lambertian	≥ 32	✘	✔	✔	✔	Point light sources, Known directions, BRDF isotropic about normal.
2007 ICCV Biswas et al. [30]	Lambertian	1	✔	✔	✘	✘	Point sources with known directions, Registered avg. 3D model required.
2007 PAMI Zhou et al. [6]	Lambertian	1	✔	✔	✘	✔	No attached shadows expected, Symmetry of faces, Bootstrap set required.
2007 IJCV Basri et al. [11]	Lambertian	15	✘	✔	✘	✔	Distant and isotropic lighting, Works for only convex objects.
2008 CVPR Alldrin et al. [17]	Non-Lambertian	≥ 102	✔	✔	✘	✔	Point sources with known directions, BRDF isotropic about normal.
Our Method	Non-Lambertian	≥ 9	✔	✔	✔	✔	Point sources with known directions.

like facial symmetry in [28]. 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 ([40]). Resolving this requires additional assumptions like symmetry of face, nose and forehead being at the same height, known lighting directions etc. or manual assistance.

Recently, shape recovery using generalized photometric stereo 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 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 [5] where graph cuts method was used to identify light source visibility and information from shadow maps were used to recover the shape.

At a contrast to most of 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 [32], where 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 [4], where radiance environment map was deduced using the ratio image technique ([42], [32]). Note that the shape recovery in [4], 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 [31], 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 [30] where the albedo estimation was made more robust using error statistics of surface normals and known illumination direction. This method requires a registered average 3D model of face and does not allow cast shadows in input but as compared to [31], it provides better shape recovery. Improving upon the idea of ideal class assumption ([32]), a generalized photometric stereo was presented in [6]. Using a bootstrap set of facial images and exploiting the subspace spanned by a set of basis object with Lambertian surface, images with novel pose and illumination were generated. Faces were assumed to symmetric and 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 [25] 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 photorealistic images of the face in novel illumination and pose. 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, 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 [38]. This method required custom built rig and more than 50 specularity free images to recover the model parameters. 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 BRDF nature was revisited in [26]. This method required input images to doubly cover the illumination direction which called for specialized data acquisition. No attempt to capture the reflectance properties of the object was made in this work.

One of the first techniques that worked with standard face databases and did not required custom data was presented in [22] where the more general Torrance-Sparrow ([51]) 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 source in order to work properly.

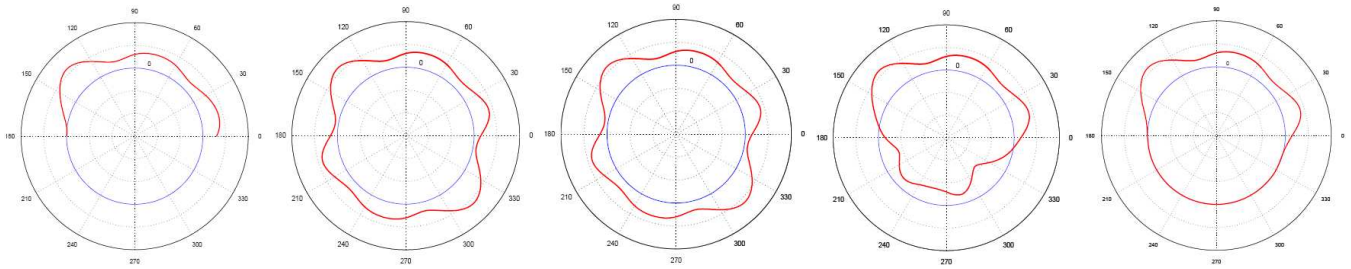
Important contribution in the field of example based shape recovery was made by [19] where objects of interest were imaged along with a reference object of known geometry (e.g. sphere). Multiple (≥ 8) cast shadows free images were used as input. This work was expanded to allow spatially varying BRDF by using multiple references in [20]. An interesting extension of this work was presented in [23] where 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. Using large amount of data obtained from a custom built rig, [24] presented a technique for pose and illumination variation using single image, that used the Morphable models to recover the 3D shape and higher-order SVD to recover the illumination subspace. This method does not explicitly make the Lambertian assumption but it requires manual initialization for the 3D model fitting.

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 human face, this work presented results with 150 images where specular component was separately captured using polarized lighting. The images were acquired from know illumination directions with no cast shadows were allowed.

Recently, [16] presented a new method for photometric reconstruction of shape 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 [17] where the need for additional constraint to recover shape from depth map isocontour was alleviated by assuming the surface to be composed of a few fundamental materials and that BRDF at each point can be approximated by a bivariate function. Results presented in this work required 102 or more images.

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 gonioreflectometer in [45], various customized hardware components and softwares were used to render face images in the movie "The Matrix Reloaded" [44] and in order to measure accurate skin reflectance while accounting for sub-surface scattering again custom built devices were employed in [46] to render high quality facial images.

It can noticed 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



(a) A semicircular function (b) Approximation of (a) using symmetric functions (c) Max of function in (b) and zero (d) Approximation of (a) using ant-symmetric functions (e) Max of function in (d) and zero

(f) ABRDF approximated with symmetric functions leads to unnatural lighting

(g) ABRDF approximated with anti-symmetric function leads to more natural lighting

Fig. 1. Symmetric and Antisymmetric ABRDF approximations.

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. Finally, the scarcity of data, paired with the nature of facial ABRDF, forces us to use only the anti-symmetric cartesian tensor components. This combination of anti-symmetric cartesian tensors and B-Splines is called Anti-Symmetric Tensor Splines.

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 neighborhood, 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 our 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 shape lower than others. This process can be repeated iteratively 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, [23] 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 ([16], [14]). Though global photo-effects like subsurface scattering and interreflection are not explicitly modeled, antisymmetric tensor splines can capture them to some extent.

4 Anti-Symmetric 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. *Note that the pose is assumed to be fixed and hence the term "ABRDF" is used to refer to a spherical function of illumination direction.*

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 Cartesian tensor (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. 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 $\mathbf{l} = (l_1 \ l_2 \ l_3)^T$ is the illumination direction, $\mathbf{x} = (x_1 \ x_2 \ x_3)^T$ is the viewing direction and ρ is the surface albedo, the Lambertian kernel is given by

$$\begin{aligned} max(\rho \cdot \mathbf{l} \cdot \mathbf{x}, 0) &= \rho \cdot max(l_1x_1 + l_2x_2 + l_3x_3, 0) \\ &= max(\sum_{k+l+m=1} T_{klm}x_1^k x_2^l x_3^m, 0) \quad (2) \end{aligned}$$

with $T_{100} = \rho \cdot l_1$, $T_{010} = \rho \cdot l_2$ and $T_{001} = \rho \cdot l_3$. 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 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.

Finally, though the higher order tensor can be more expressive, they can be perceived to be more sensitive to noise due to their ability to model high frequency details. In contrast, the lower order tensors are incapable of modeling high frequency information but arguably are more robust to noise. Since it is impossible to discriminate between high frequency detail and noise in the data, it is reasonable to say that the high order tensors possess higher noise sensitivity. Thus, any approximation task must strike a balance between the high frequency data fidelity and the noise sensitivity.

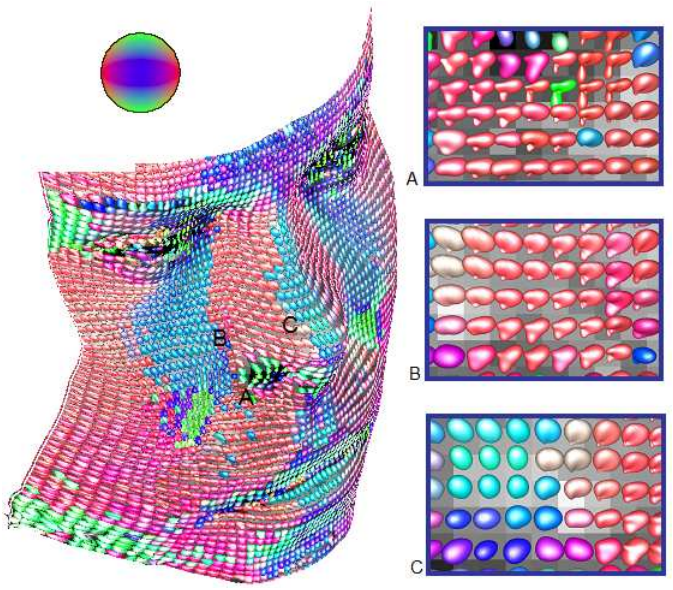


Fig. 2. Recovered ABRDFs for a human face. Complex shapes of ABRDFs in various regions of the face can be readily noted. *(This image is best viewed in color.)*

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 high noise sensitivity of higher order tensors, it might be reasonable to enforce smoothness across the field of spherical functions. We accomplish this by combining Cartesian tensor basis at each pixel with the B-Spline basis ([47]) across the lattice of spherical functions.

We define a tensor splines as a B-spline of multilinear functions of any order. In a tensor spline, the multilinear functions are weighted by the B-spline basis $N_{i,k+1}$, where

$$N_{i,1} = \begin{cases} 1 & \text{if } t_i \le t < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

and

$$N_{i,k}(t) = N_{i,k-1}(t)\frac{t - t_i}{t_{i+k-1} - t_i} + N_{i+1,k-1}(t)\frac{t_{i+k} - t}{t_{i+k} - t_{i+1}}, \quad (4)$$

The $N_{i,k+1}(t)$ are polynomials of degree k, associated with n+k+2 monotonically increasing numbers called "knots" $(t_{-k}, t_{-k+1}, ..., t_{n+1})$.

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

$$S(\mathbf{t}, \mathbf{v}) = \sum_{a=1...p} (\prod_{i_a} N_{i_a,k+1}(t_1 \dots t_p))T_{i_1 \dots i_p}(\mathbf{v}) \quad (5)$$

where $\mathbf{t} = (t_1 \dots t_p)$ is the index into the lattice, $\mathbf{v} = (v_1 \ v_2 \ v_3)^T$ is a unit vector, and $T_{i_1 \dots i_p}(\mathbf{v})$ is given by Eq. 1. In the 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. 5 is quiet general as it can be used to estimate spherical function field defined over

Fig. 3. 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.

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 while modeling the ABRDFs. Note that here the lattice is 2-dimensional and the assumption of locally 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.

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 anti-symmetric or the symmetric components. Note that since most of the time we are interested in ABRDFs' behavior on the frontal hemisphere, both symmetric and anti-symmetric components provide the same representation power. Their behavior only become pertinent when the illumination direction is exactly perpendicular to the pose direction, and this is where use of anti-symmetric components is advantageous.

This has been explained via an 2D example in Fig. 1 where the Fig. 1(a) shows a semicircular function. The blue circle in the figure is considered to be the zero value. Fig. 1(b) and 1(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 (Fig. 1(a)) is such that it has positive value at one of these antipodal points and near zero value at the other, 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 Fig. 1(c) and 1(e).

The behavior of most facial ABRDFs is similar to the function in Fig. 1(a). This is because if a pixel has high intensity value when lit from 0°, most of the time it would have a dark 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 the both 0° and 180° and would lead to visually significant artifacts (unnatural lighting) in the images (Fig. 1(f)). On the other hand, in practise, use of anti-symmetric function does not cause visually significant artifacts (Fig. 1(g)). To summarize, even though both, anti-symmetric and symmetric functions, introduce artifacts near 0° and 180° directions, the artifacts created by 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 anti-symmetric tensors can be written as

$$S(\mathbf{t}, \mathbf{v}) = \sum_{i,j} N_{i,4}(t_x) N_{j,4}(t_y) T_{i,j}(\mathbf{v}) \tag{6}$$

where vectors $i, j, \mathbf{t}$ 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, $q = 1 \ldots Q$) 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 squeezes the L_2 distance between the model and the

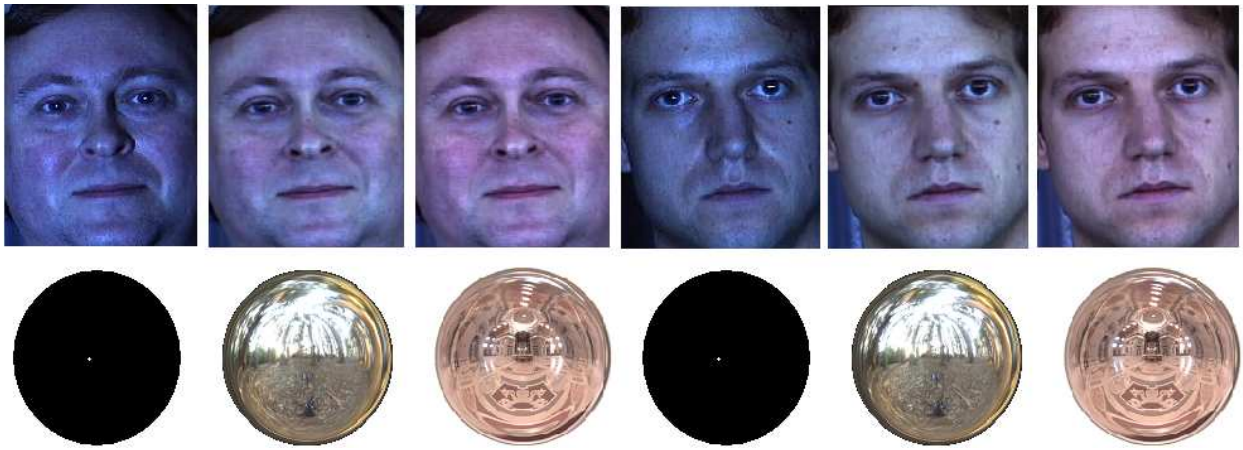


Fig. 4. Images relit with complex lighting. 1^{st} image of both subjects is lit by a point source while the next two are lit by Eucalyptus Grove and St. Peter's Basilica light probes respectively. Light probes are provided below the facial images. *(This image is best viewed in color.)*

given data,

$$\mathbf{E}(T_{ijklm}) = \sum_{q=1}^{Q}\sum_{t_x,t_y}(\sum_{i,j} N_{i,4}(t_x)N_{j,4}(t_y)T_{i,j}(\mathbf{v}_q) - I_q(t_x,t_y))^2 = \sum_{q=1}^{Q}\sum_{t_x,t_y}(\sum_{i,j} N_{i,4}(t_x)N_{j,4}(t_y)\sum_{k+l+m=n} T_{ijklm}v_{q1}^k v_{q2}^l v_{q3}^m - I_q(t_x,t_y))^2 \quad (7)$$

where t_x, t_y run through the lattice of the given images and the tensor order n is an odd integer. The minimization of Eq. 7 is done with respect to the unknown tensor coefficients $T_{i,j,k,l,m}$ that correspond to the control tensors $T_{i,j}(\mathbf{v}_n)$.

To keep the formulation simple we use an uniform grid of knots in both lattice coordinates. If the pixel lattice size in the given images is $M \times M$, there are $(M+2)\times(M+2)$ control tensors. The number of unknown coefficients at each control tensor is governed by the order of the tensor. In the case of 1^{st} order tensors, there are $3(M+2)^2$ unknowns, in the case of 3^{rd} order tensor, $10(M+2)^2$ and in the case of 5^{th} order tensor, there are $21(M+2)^2$ unknowns.

We recover the unknowns in Eq. 7 using the non-linear conjugate gradient method with control tensor coefficient field initialized to unit vectors. This technique can be efficiently implemented because we have obtained the closed form of the derivative of the objective function with respect to the unknown coefficients as

$$\partial\mathbf{E}(T_{ijklm})/\partial T_{ijklm} = 2\cdot(\sum_{q=1}^{Q}\sum_{t_x,t_y}(\sum_{i,j} N_{i,4}(t_x)N_{j,4}(t_y)T_{i,j}(\mathbf{v}_q) - I_q(t_x,t_y))) \times(\sum_{q=1}^{Q}\sum_{t_x,t_y}(N_{i,4}(t_x)N_{j,4}(t_y)v_{q1}^k v_{q2}^l v_{q3}^m)). \quad (8)$$

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. 5. Possible negative values obtained in Eq. 5 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. 5 on a more dense sampling lattice since the tensor spline is a continuous function.

5 Continuous Mixture of Single-Lobed Functions

In order to quantitatively validate whether the tensor spline provide a good enough approximation of ABRDF field or not, 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 a continuous mixture ([48]) of single-lobed spherical functions. A continuous mixture is characterized by a kernel function and a mixing density and for the spherical domain it can be represented as

$$B(\mathbf{v}) = \int_{S_2} f(\mathbf{u})k(\mathbf{u},\mathbf{v})d\mathbf{u} \quad (9)$$

where $\mathbf{u}$ and $\mathbf{v}$ are directions and the integration is over the sphere.

Of the various choices for singled lobed spherical functions that can be used as kernel function $k(\mathbf{u},\mathbf{v})$, we pick $k(\mathbf{u},\mathbf{v}) = e^{-\mathbf{u}\cdot\mathbf{v}} - 1$ primarily due to three reasons. Firstly, it has a single peak, secondly, $k(\mathbf{u},\mathbf{v}) = 0$ for all $\mathbf{v}$ such that $\mathbf{v}\cdot\mathbf{u} = 0$ (because if the viewing and illumination directions are perpendicular we expect zero intensity) and thirdly, it leads to a closed form for the continuous mixture which facilitates efficient implementation. Note that the first two properties are also satisfied by the Lambertian kernel.

Since the kernel function is parameterized by directions, we use a mixing density, $f(\mathbf{u})$, which is also

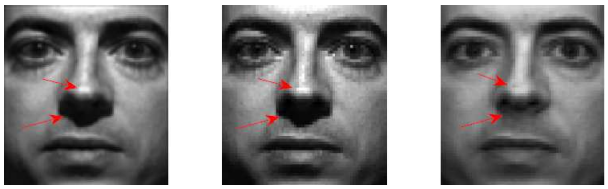


Fig. 5. The 1^{st} image is generated using tensor splines, 2^{nd} image is the ground truth and the 3^{rd} is generated using Lambertian model. Cast shadows and specularities are much more realistically render using tensor splines than the Lambertian model.

defined on the space of directions. Spherical analog of the Gaussian density, von-Mises Fisher density is the natural choice here and it is expressed as

$$f(\mathbf{x}|\kappa,\mu) = \frac{\kappa}{4\pi sinh(\kappa)} \cdot e^{\kappa\mu\cdot\mathbf{v}}, \tag{10}$$

where μ is a unit vector defining the mean orientation and κ is a scalar governing the concentration of the distribution. Substituting our choices for kernel and mixing density in Eq. 9, we obtain the following expression:

$$B(\mathbf{v}) = \int_{S_2} \frac{\kappa}{4\pi sinh(\kappa)} \cdot e^{\kappa\mu\cdot\mathbf{u}}(e^{-\mathbf{u}\cdot\mathbf{v}} - 1)d\mathbf{u}. \tag{11}$$

Here we make the *important* observation that this integral is the Laplace transform of the von Mises-Fisher distribution, which we have analytically computed to be

$$B(\mathbf{v}) = \frac{\kappa sinh(\|\kappa\mu - \mathbf{v}\|)}{sinh(\kappa)\|\kappa\mu - \mathbf{v}\|} - 1. \tag{12}$$

However, since the expression in Eq. 12 is a convolution of two single lobed function, it itself is single lobed. Therefore, to get a model with possibly multiple lobes, we propose to use a finite mixture of von Mises-Fisher distributions as the mixing density, which leads to an alternate definition of mixing density as $f(\mathbf{u}) = \sum_i w_i f(\mathbf{u}|\kappa,\mu_i)$, where w_i are the mixture weights and $\mu_\mathbf{i}$ are the various direction along which the component von Mises-Fisher distributions are oriented. Using this mixture of von Mises-Fisher distributions we obtain the following expression for the continuous mixture

$$B(\mathbf{v}) = \sum_i w_i \left(\frac{\kappa sinh(\|\kappa\mu_i - \mathbf{v}\|)}{sinh(\kappa)\|\kappa\mu_i - \mathbf{v}\|} - 1\right). \tag{13}$$

We must emphasize that although $f(\mathbf{u})$ has the form of a discrete mixture, the approximating function $B(\mathbf{v})$ is still a continuous mixture of the single-lobed kernel functions $k(\mathbf{u},\mathbf{v})$.

The task of estimating ABRDFs using this continuous mixture 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 Eq. 13 for every $\mathbf{v}_n$ and μ_i. M is the number of μ_i picked in the model. The unknown weights (Eq. 13) 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 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 [49].

Note that this model would generally have a very large number of unknowns (depending on the chosen resolution while picking μ_i), governed by number of 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 not considered a drawback.

6 Recovering Shape from 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 ([51]), Phong([50]) etc., which explicitly assume a role for surface normal in their formulae, tensor splines make no such assumption, which allows spatially varying and a accurate approximation of the ABRDFs, but also makes the recovery of 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 does not change drastically in small neighborhoods (3×3 pixels) and mathematically robust as tensor splines ensures that that coefficient 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 surface normal at one of these pixels is known, surface normal at the other pixel can be derived using the rotating R.

For a given internal pixel (x,y) in the image, there are eight immediate neighbors. If surface normal at (x,y) is inferred as described above, it would get eight suggestions for possible surface normals (assuming that the surface normals for 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 of computing the surface normal is this manner. Firstly, being an aggregate statistic, geodesic mean is more robust to noise than individual suggestions. Secondly and more importantly, the weighted nature of the mean ensures that suggestions, which originate from neighbors whose ABRDFs are too 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 is weak, e.g. shadow edges.

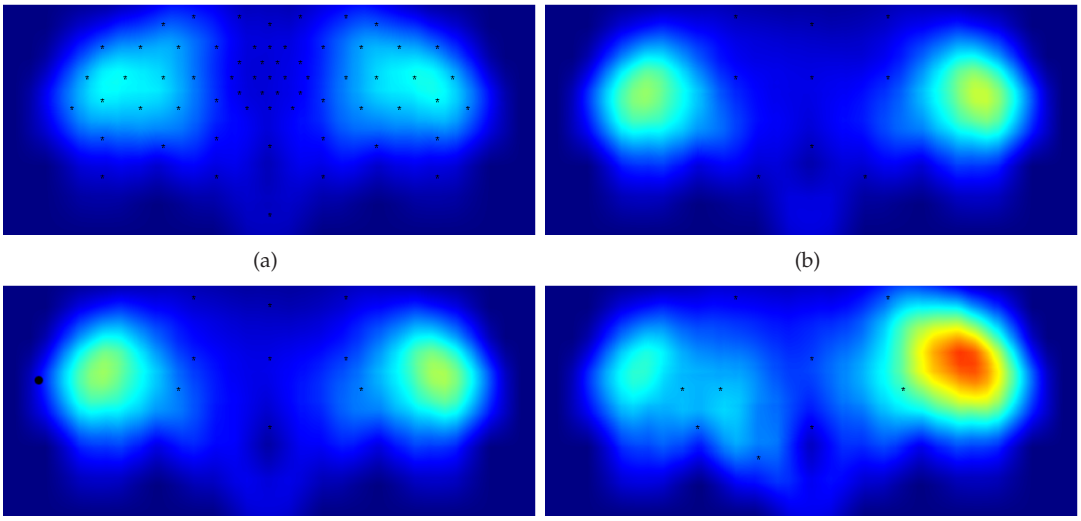


Fig. 6. Distribution of errors as the configuration of input changes. X-axis represents azimuth and Y-axis represents elevation angles. Hotter colors show larger errors. The black dots represent the exact directions of illumination in images used as input. *(This image is best viewed in color.)*

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 the maxima. Initialization is followed by weighted geodesic mean computation which provides us with a robust estimate of the surface normals. The process of mean computation can be 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 neighbor 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

$$E(R) = \sum_{\mathbf{v} \in S^2} (w_1^T B(\mathbf{v}) - w_2^T B(R \cdot \mathbf{v}))^2. \tag{14}$$

such that

$$R^T R = I. \tag{15}$$

where unit vector $\mathbf{v}$ is obtained by some uniform sampling of the sphere, 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 inefficient 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 we seek a $\delta\mathbf{v}$ such that

$$T_1(\mathbf{v}) = T_2(\mathbf{v} + \delta\mathbf{v}). \tag{16}$$

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

$$T_1(\mathbf{v}) = T_2(\mathbf{v}) + \nabla T_2(\mathbf{v})^T \delta\mathbf{v}. \tag{17}$$

As we expect

$$L \cdot \mathbf{v} = \mathbf{v} + \delta\mathbf{v}, \tag{18}$$

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

$$T_1(\mathbf{v}) - T_2(\mathbf{v}) + \nabla T_2(\mathbf{v})^T \mathbf{v} = \nabla T_2(\mathbf{v})^T L\mathbf{v}, \tag{19}$$

which leads to the linear system

$$Ax = B, \tag{20}$$

where 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, 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 v_i are the unit vector obtained from some 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

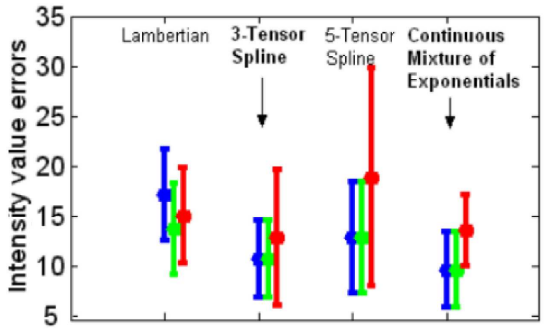


Fig. 7. Per pixel intensity error comparison. The blue color show errors on 1st and 2nd subset combined which contain lighting directions, ϕ, smaller than 25°, the green color shows the 3rd subset with $25^\circ < \phi < 50^\circ$ and the red color shows the 4th subset with $50^\circ < \phi < 70^\circ$. *(This image is best viewed in color.)*

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 given by

$$\mathbf{n} = argmin_\mu \sum_{i=1}^{p} d(\frac{1}{||T_p - T||^2}\mathbf{n}_p, \mu), \tag{21}$$

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 Karcher mean and can be computed using the following iterative scheme –

$$\mu \rightarrow exp_\mu(\epsilon\nu) \tag{22}$$

$$\nu = (1/n) \sum_{i=1}^{p} exp_\mu^{-1}\mathbf{n_p} \tag{23}$$

where exp, the exponential map, is given as

$$exp_\mu(\epsilon\nu) = \cos(|\epsilon\nu|)\mu + \sin(|\epsilon\nu|)(\nu/|\nu|) \tag{24}$$

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

$$exp_\mu^{-1}(n_p) = u \cos^{-1}(\langle \mu, n_p \rangle)/\sqrt{(\langle u, u \rangle)} \tag{25}$$

where

$$u = n_p - \langle n_p, \mu \rangle \mu, \tag{26}$$

and ϵ is the iteration step size. For more details on computing means on manifolds see [53] and references therein.

6.3 Shape Recovery

Once the normal field has been computed, we use one of the standard techniques ([52]) to recover surface. If $z(x, y)$ defines the surface, normal a location (x, y) is given by $(z_x \; z_y \; -1)^T$ where z_x and z_y denote the partial derivative 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

$$z_x = -n_x/n_z \tag{27}$$

$$z_y = -n_y/n_z. \tag{28}$$

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

$$n_z z(x+1, y) - n_z z(x, y) = n_x \tag{29}$$

$$n_z z(x, y+1) - n_z z(x, y) = n_y, \tag{30}$$

which provide a linear relation between the surface values at the grid point and the known surface normals. Surface can now be recovered by solving an overdetermined system of linear equations. At the boundary points, the above formulation is not valid and the surface is recovered by solving the following equation, obtained by eliminating n_z above

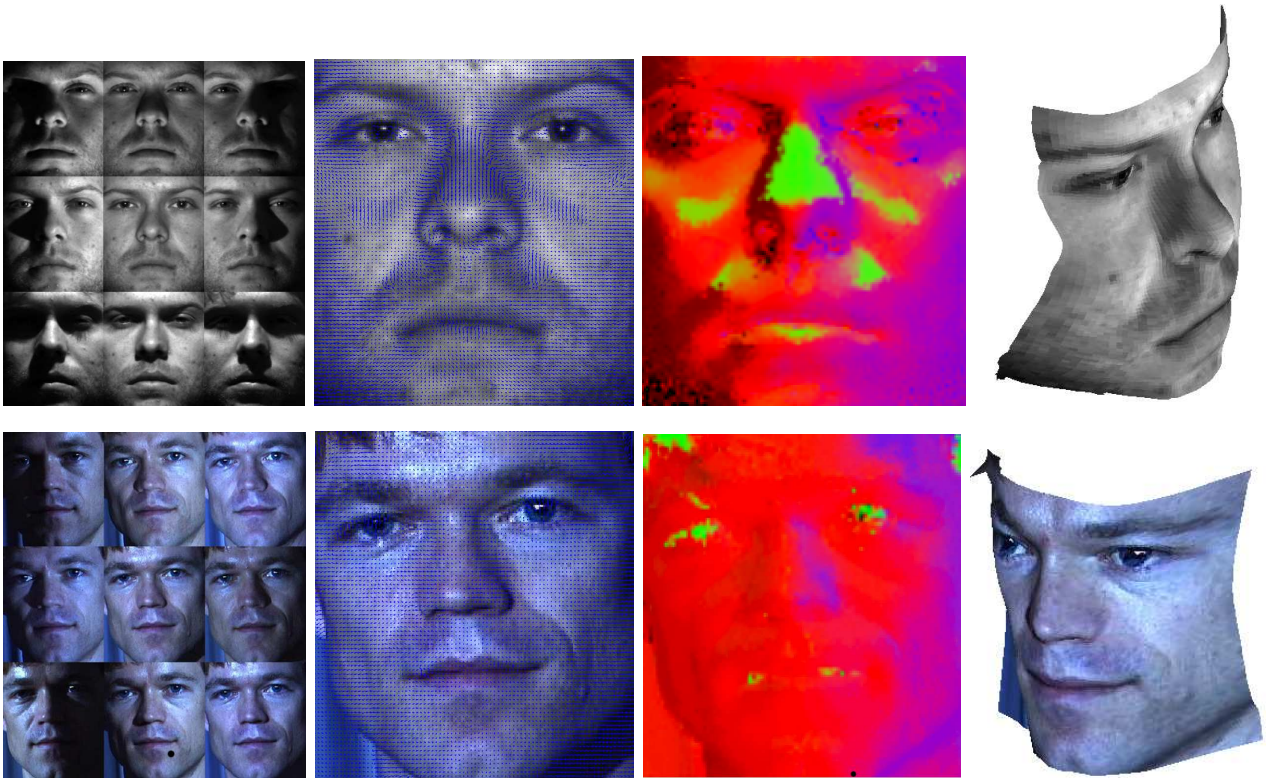
$$n_x z(x, y) - n_x z(x, y+1) = n_y z(x+1, y) - n_y z(x, y). \tag{31}$$

6.4 Novel Pose Relighting

With the facial shape at hand, novel poses can be rendered by simply changing the viewpoints. But generating novel illumination conditions in the novel pose is not trivial as ABRDFs estimated from a different pose cannot be directly use. If the ABRDF field was estimated in pose P_1 and if we wish to generate image with novel illumination in a new pose P_2, we have 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 sampling the ABRDF field in 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 specular component by using polarized lighting (e.g. [14], [25]), which required specialized data acquisition, or by assuming a parametric form for the specular component of lighting (e.g. [22]).

Our Cartesian tensor representation for ABRDF does not discriminate against specularities and estimates the ABRDF as best possible from the available intensity values. Thus it should possible to recover and manipulate specular component separately, but at this stage, we make the assumption that specularities do not change drastically across facial poses. The validity of this assumption is supported by the results presented in the next section.



(e) 9 input images (f) Quiver plot of the normal field (g) Color encoded normals (R-x, G-z, B-y) (h) Recovered shape in novel pose

Fig. 8. Recovered shapes from 9 images. *(This image is best viewed in color.)*

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 quiet 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 splines based smoothness constraint, at most 10 coefficients can be recovered and hence our central results use bi-cubic 3^{rd} order tensor splines.

The experiments were carried out on Extended Yale B [43] (28 subjects, in 9 poses and 64 illumination conditions) and CMU PIE [54] (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.

7.1 Relighting faces

We begin by noting that anti-symmetric tensor splines can capture non-trivial shapes of facial ABRDFs. In Fig. 2 we show the ABRDF field of a subject from Extended Yale B database estimated using 9 images. 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 shapes because these ABRDFs have to accommodate shadows. The spherical functions in the image have been color coded based on maximal value direction. Mapping of directions to colors is provided in the lower right corner.

Next we present results for relighting of faces in novel illumination direction. In Fig. 3 four different subjects lit in various novel point source illumination are depicted. For the first two rows the illumination direction varies across the azimuth angle while in the next two rows variation is in the elevation angle. It can 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.

Starting with 9 images, our technique estimates the entire ABRDF field and thus images lit in fairly complex lighting conditions can be rendered. In Fig. 4 we present such results for two subjects from the CMU PIE database. Below each image is its lighting condition. The first image of both subjects is one of the nine input images taken to estimate their ABRDF fields. The next two images for each subjects are lit by light probes ([55]) 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

Fig. 9. Detailed pose variation with texture-less in upper right and depth-map in lower right. Accurate renderings even in extreme poses can be noticed.

lit images where the weights were determined using the light probes. We used 2500 samples of the light probe to render these images.

In Fig. 5 we provide a qualitative comparison between our method and the Lambertian model. The first face image in Fig. 5 is render using tensor splines, the second is the ground truth and the third image is rendered using the Lambertian model. It can be readily noted that image obtained using our method is much more closer to the ground truth than the one rendered using Lambertian model. The arrows show the locations of important differences – cast shadows and specularities.

The next two experiments try to quantitatively capture the performance of our method. First, we explore the impact of the input data on the estimation when the tensor order is fixed (3^{rd} order in this case). For this we use Extended Yale B dataset as it provides ground truth for 64 directions. 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 image used as input. Next, errors were computed similarly for 3 other cases where only 9 images were used as input to our method but in different configurations. Two of the cases had images with illumination directions uniformly distributed in front of the face while one had images with direction biased toward one side.

To visualize the distribution of the obtained errors, we color coded them with large error in hotter colors and smaller error in cooler colors and plotted them as a continuous images in Fig. 6. The X axes of these images show azimuth angle varying from -130° to 130° (from leftmost black dot to rightmost) and Y axes show elevation angle varying form -40° to 90° (from topmost black dot to bottom most). The black dots in these images show the exact direction of illumination in the images used as input. It can be readily noted that when all 64 images are used as input Fig. 6(a), the error is the least. For the 9 image case, Fig. 6(b) and 6(c), where the illumination directions in input images are uniformly distributed, the error is more than 6(a) but notedly less than the case when distribution is skewed in one direction, 6(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 affected too much by the input image distribution.

Next we present a quantitative comparison of our method with Lambertian model and the validation model presented in Section 5. A natural question that arises is that why should an order 3 Cartesian tensor be suitable for estimating the ABRDF? To answer this question, we computed the average intensity error per pixel over all 48 subjects in 64 illumination directions of Extended Yale B dataset using the Lambertian model, 3^{rd} order tensor splines, 5^{th} order tensor spline and continuous mixture of exponentials. All 64 illumination directions were used for the continuous mixture model (on account of large number of unknowns) while for the other three only 9 images (configuration shown in Fig. 6(b)) were used. We set the μ_i values required for the continuous mixture model using a dense sampling (642 directions) of the unit sphere obtained by the 4^{th}-order tessellation of the icosahedron. We have presented

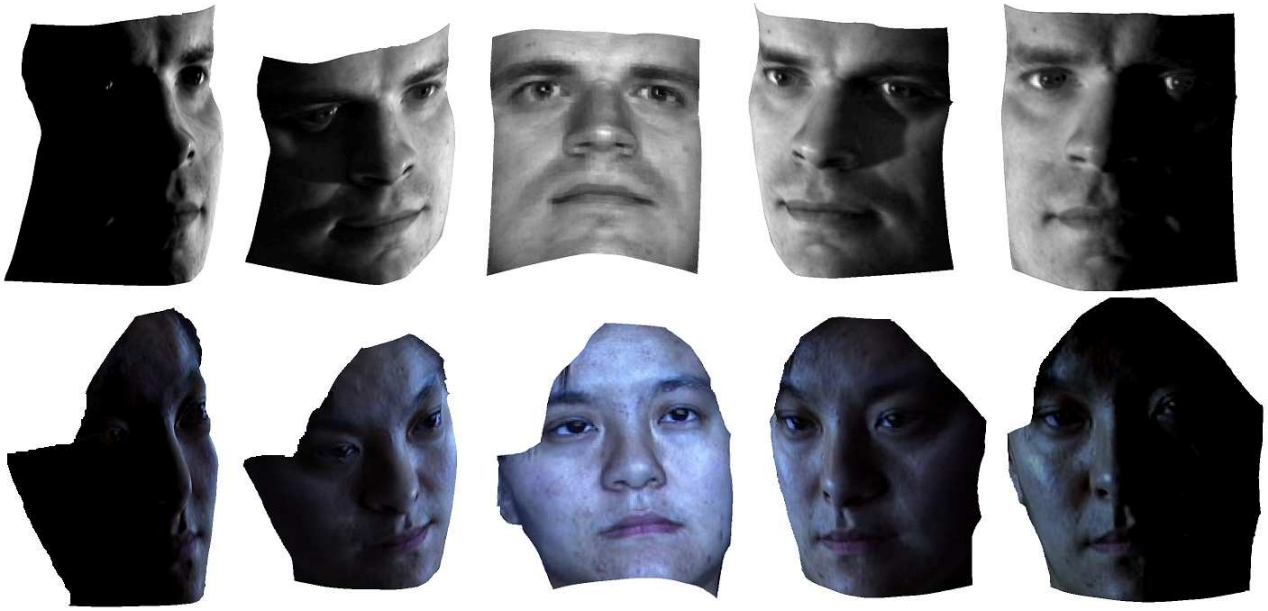


Fig. 10. Simultaneous pose and illumination variation. *(This image is best viewed in color.)*

results in Fig. 7 shattered along the standard subsets (subset 4 in red, 3 in green and 1 + 2 in blue) of Extended Yale B database. As expected, error for the subset with extreme lighting (subset 4) is more than other other set for all method. More importantly, even with considerably large amount of data and very flexible estimation model, the errors obtained from continuous mixture of exponentials method is quite similar to those obtained from 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 is suitable enough. The 3^{rd} order tensor spline outperforms the Lambertian model and even the 5^{th} order tensor spline, which suggests possible over-fitting.

7.2 Estimating Shape

All the results presented till now assumed a fixed pose but using the technique presented in Section 6 we can simultaneously vary illumination and pose of a face. Fig. 6.3 summarizes the results produced by our shape recovery algorithm for one subject each from Extended Yale B and CMU PIE databases. The first column shows the 9 input images, the second column shows the quiver plot of the estimated normal field (zoom in to see details), the third column present the surface normal information in a color coded form (x components of normal field are mapped to red channel, y components to blue and z components to green channel) and the fourth column shows the recovered shape in a novel pose. For the case of color image, shape estimation was carried out using only the luminance component. In both the cases occlusion of appropriate regions of the face due to pose change can be noted from the images in the fourth column.

In Fig. 9 we present more detailed results for pose variation with a fixed illumination. The 3 rows of images show a subject from Extended Yale B in different poses ranging from right profile to left profile as we go from left to right and viewpoint varying from below to face to above the face as we go from top to bottom. Note that the ABRDF filed for this subject recovered using just 9 images under the illumination configuration shown in Fig. 6(b). The Recovered shape for the same subject, rendered with constant albedo and specularities is also presented towards the right end of the figure. This allows finer details of the shape to be shown without any texture to bias the observer. And finally, to the lower right of the figure is the height map for the same subject. It can be noted that our shape recovery algorithm can produce good results without making simplifying Lambertian assumption.

Finally, we present results when both pose and illumination conditions are simultaneously varied. In Fig. 10 one subject each from CMU PIE and Extended Yale B database are shown in various poses and illumination conditions. The ABRDF fields for both cases were recovered using 9 images and shape for the color image 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 application of facial image analysis. It is generally defined as – given a database of facial images of various people, called gallery, identify the person in a novel test image, called probe, as one the people present in the database. The

degree of difficulty of this problem increases as the difference between the probe and the gallery images increases. This difference could be in illumination, occlusion, expression, pose or any combination of these.

In recent times, illumination invariant face recognition has attracted particular interest due advances in our understanding of reflectance modeling. Here we present a comparative study of illumination invariant face recognition. When using tensor splines, we assume that for each subject 9 gallery images with known illumination directions are available. From these images we compute the ABRDF field and generate images with novel illumination for a dense sampling of directions. This step exapnds our collection of 9 gallery images to any desired size. The probe image is then matched to all the images in the database and the subject with the closest matching image is assumed to be the correct identity of the probe image.

We have used Extended Yale B data for this experiment primarily because most of the existing methods have presented face recognition results on the same database. As mentioned before, the database is divided into 4 subsets with the lighting getting more and more extreme as we go from subset 1 to 4 and thus, the difficulty to classify images from these subsets also increases. The obtained recognition error rate are reported in Table 2. We have also presented results reported by existing methods and their respective references are listed next to the method name. Results for the first seven techniques were taken from [7] and rest were taken from 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. 6(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 is the number of required input images.

Method	N	Subset 1&2	Subset 3	Subset 4	Total
Correlation [3]	4	0.0	23.3	73.6	29.1
Eigenfaces [8]	6	0.0	25.8	75.7	30.4
Linear subspace [2]	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 [43]	9	0.0	0.0	2.8	0.8
3D SH [7]	1	0.0	0.0	2.8	0.8
Harmonic (SFS) [1]	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 anti-symmetric tensor splines. This method can be used to synthesize photo-realistic images of a faces in novel poses and illumination conditions. Our method require as few as 9 images with known illumination directions in order to fully recover the model parameters and the accuracy of the approximation improves as more data is made available. The output of this technique is a continuously varying field of ABRDFs which can be used for, besides relighting and shape recovery, meaningfully up sampling the image and succinctly describing the ABRDF field of the face. It should be noted that our method does not require the input images to be free of attached or cast shadows.

As compared to the ideal solution described in the introduction, we work with multiple images lit with point sources in known illumination direction. Our future work would focus on recovering the ABRDF field without requiring multiple images with known illumination. Further, while relighting novel poses, we assume that the specular component of the ABRDF remains the same which is not precise. As we recovered the surface normals from the non-trivial shapes of ABRDFs, we believe that the specular component can also be recovered, but that is yet to be explored. Lastly, in order to avoid over fitting noisy and sparse data, we built anti-symmetric tensor spline with inherent local but uniform smoothness constraints, but ideally one would like to have more control over this and it is something that we also wish to explore in future. Note that these assumptions are loosening of a few of the constraints we set out to satisfy while looking for an "ideal" solution but almost all existing techniques make similar or more restrictive assumptions (Table. 1).

In conclusion, anti-symmetric tensor splines provide a useful framework for analysis and modeling of illumination and pose variation of facial images with various possible applications like relighting, pose change, face recognition, up-samplings, compression etc. It also shows that the analysis of shape and reflectance in a collective manner through ABRDFs seems promising as an alternate 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 [29]. This research was in part funded by the University of Florida Alumni Fellowships to Ritwik Kumar and Angelos Barmpoutis.

References

[1] W.A.P. Smith, E. R. Hancock, "Facial Shape-from-shading and Recognition Using Principal Geodesic Analysis and Robust Statistics", IJCV, 76:7191, 2008.
[2] A. U. Batur and M. Hayes, "Linear subspaces for illumination robust face recognition", CVPR, 2:296301, 2001.
[3] T. Brunelli, R. Poggio, "Face recognition: features versus templates", PAMI, 15(10):10421052, Oct 1993.
[4] Z. Wen, Z. Liu and T. Huang,"Face Relighting with Radiance Environment Maps", CVPR, page 158-165, 2003.
[5] M. Chandraker, S. Agarwal and D. Kriegman, "ShadowCuts: Photometric stereo with shadows", CVPR, 2007.

[6] 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 29(2):230-245, February 2007.
[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, pages 995999, 1994.
[9] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces", SIGGRAPH, pp. 187194, 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 and P. Belhumeur, "Reflectance Sharing: Predicting Appearance from a Sparse Set of Images of a Known Shape", PAMI, 28(8), August 2006.
[15] C. Hernndez, G. Vogiatzis, G. Brostow, B. Stenger and R. Cipolla, "Non-rigid Photometric Stereo with Colored Lights", ICCV, 2007.
[16] N. Alldrin and D. Kriegman, "Shape from Varying Illumination and Viewpoint", ICCV, 2007.
[17] Neil Alldrin, Todd Zickler, and David Kriegman, "Photometric Stereo With Non-Parametric and Spatially-Varying Reflectance", CVPR, 2008.
[18] W. M. Silver, "Determining Shape and Reflectance Using Multiple Images", Masters thesis, MIT, 1980.
[19] A. Hertzmann and S. M. Seitz, "Shape and materials by example: A photometric stereo approach", CVPR, 2003.
[20] A. Hertzmann and S. M. Seitz, "Example-based photometric stereo: Shape reconstruction with general, varying brdfs", PAMI, 27(8):12541264, 2005.
[21] A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman, "Illumination-Based Image Synthesis: Creating Novel Images of Human Faces Under Differing Pose and Lighting", Proceedings of the IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes, page 47, 1999.
[22] A. S. Georghiages, "ecovering 3-D Shape and Reflectance From a Small Number of Photographs", Eurographics Symposium on Rendering, June 2003, pp. 230-240, 315.
[23] D.B. Goldman, B. Curless, A. Hertzmann, S. M. Seitz, "Shape and Spatially-Varying BRDFs from Photometric Stereo", ICCV, pp. 341-348, 2005.
[24] J. Lee, B. Moghaddam, H. Pfister and R. Machiraju, "A Bilinear Illumination Model for Robust Face Recognition", ICCV, pp 1177-1184, 2005.
[25] P. Debevec, T. Hawkins, C. Tchou, H. Duiker, W. Sarokin, and M. Sagar, "Acquiring the Reflectance Field of a Human Face", SIGGRAPH, pp. 145-156, 2000.
[26] S. Magda, T. Zickler, D. Kriegman and P. Belhumeur, "Beyond Lambert: Reconstructing Surfaces with Arbitrary BRDFs", ICCV, 2001.
[27] Z. Yue, W. Zhao and R. Chellappa, "Pose-Encoded Spherical Harmonics for Face Recognition and Synthesis Using a Single Image", EURASIP Journal on Applied Signal Processing, 2008.
[28] W. Zhao and R. Chellappa,"Symmetric Shape from Shading Using Self-Ratio Image", IJCV, Vol. 45, pp. 55-75, 2001.
[29] A. Barmpoutis, R. Kumar, B. C. Vemuri and A. Banerjee,"Beyond the Lambertian Assumption: A generative model for Apparent BRDF fields of Faces using Anti-Symmetric Tensor Splines", CVPR 08, 2008.
[30] S. Biswas, G. Aggarwal and R. Chellappa, "Robust Estimation of Albedo for Illumination-invariant Matching and Shape Recovery", ICCV, 2007.
[31] K. C. Lee and B. Moghaddam, "A practical face relighting method for directional lighting normalization", International Workshop on Analysis and Modeling of Faces and Gestures, 2005.
[32] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations", PAMI, 23(2):129-139, 2001.
[33] R. Gross, I. Matthews and S. Baker, "Appearance-Based Face Recognition and Light-Fields.", PAMI, 26(4):449-465, 2004.
[34] H. Chan and W. W. Bledsoe, "A Man-Machine Facial Recognition System - Some Preliminary Results", Panoramic Research Inc, Palo Alto, 1965.
[35] A Shashua,"On Photometric Issues in 3D Visual Recognition from a Single 2D Image" IJCV, vol. 21, pp. 99-122, 1997.
[36] A.L. Yuille, D. Snow, R. Epstein and P.N. Belhumeur, "Determining Generative Models of Objects under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability", IJCV, vol. 35, pp. 203-222, 1999.
[37] P.N. Belhumeur and D.J. Kriegman, "What Is the Set of Images of an Object under All Possible Illumination Conditions?" IJCV, vol. 28, pp. 245-260, 1998.
[38] T. Malzbender, D. Gelb, and H.Wolters. "Polynomial Texture Maps", SIGGRAPH, pages 519528, 2001.
[39] R. Ramamoothi and P. Hanrahan, "On the Relationship between Radiance and Irradiance: Determining the Illumination from Images of a Convex Lambertian Object," Journal of Optical Society of America, vol. 18, pp. 2448-2459, 2001.
[40] P. Belhumeur, D. Kriegman, A. Yuille, "The Bas-Relief Ambiguity", IJCV, 35(1), 1999, pp. 33-44
[41] 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 (APGV) 2008, August 2008, pp 135 - 142.
[42] Z. Liu, Y. Shan, and Z. Zhang, "Expressive expression mapping with ratio images", SIGGRAPH, pages 271276, 2001.
[43] K. Lee, J. Ho and D. Kriegman, "Acquiring Linear Subspaces for Face Recognition under Variable Lighting", PAMI, 27(5): 684-698, 2005.
[44] G. Borshukov and J.P. Lewis, "Realistic Human Face Rendering for "The Matrix Reloaded", Sketches and Applications Program, SIGGRAPH, 2003.
[45] S. C. Foo. "A gonioreflectometer for measuring the bidirectional reflectance of material for use in illumination computation", Masters thesis, Cornell University, Ithaca, 1997.
[46] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross, "Analysis of Human Faces using a Measurement-Based Skin Reflectance Model", SIGGRAPH, Boston, 2006.
[47] C. de Boor, "On Calculating with B-Splines", Journal of Approximation Theory, 6:5062, 1972.
[48] B. Jian and B. C. Vemuri, "Multi-fiber reconstruction from Diffusion MRI using Mixture of Wisharts" IPMI, 384395, 2007.
[49] C. Lawson and R. J. Hanson, "Solving Least squares problems", Prentice-Hall, Englewood Cliffs, 1974.
[50] B.T. Phong, "Illumination for computer generated images", Communications of the ACM, 18(6):311317, June 1975.
[51] K.E. Torrance and E.M. Sparrow, "Theory for off-specular reflection from roughened surfaces", J. Opt. Soc. Am., 57:11051114, 1967.
[52] B.K.P. Horn, "Robot Vision", MIT Press, 1986.
[53] A. Srivastava, I. Jermyn and S. Joshi, "Riemannian Analysis of Probability Density Functions with Applications in Vision", CVPR, June, 2007.
[54] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces", Technical Report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, January 2001.
[55] Paul Debevec, "Rendering Synthetic Objects Into Real Scenes: Bridging Traditional and Image-Based Graphics With Global Illumination and High Dynamic Range Photography", Proceedings of SIGGRAPH 98, pp. 189-198.