Eigenbubbles: An Enhanced Apparent BRDF Representation

Ritwik Kumar
SEAS, Harvard University
rkkumar@seas.harvard.edu

Baba C. Vemuri, Arunava Banerjee
Dept. of CISE, University of Florida
{*vemuri,arunava*}*@cise.ufl.edu*

Abstract

In this paper we address the problem of relighting faces in presence of cast shadows and specularities. We present a solution to this problem by capturing the spatially varying Apparent Bidirectional Reflectance Functions (ABRDF) fields of human faces using Spline Modulated Spherical Harmonics and representing them using a few salient spherical functions called Eigenbubbles. Through extensive experiments on the Extended Yale B and the CMU PIE benchmark datasets we demonstrate that the proposed method clearly outperforms the state-of-the-art techniques in synthesized image quality. Furthermore, we show that our framework allows for ABDRF field compression and can also be used to enhance performance of face recognition algorithms.

1 Prologue

Over the course of last decade, various methods have been proposed to tackle the problem of face relighting. Even though high quality results can be produced given large amount of resources (e.g. [6]), the problem is still largely unsolved when a more realistic constrained resource scenario is considered. The existing solutions to the constrained version of the problem - where only few images of each face are available and custom built specialized devices are not accessible - largely work with the Lambertian assumption([3, 10, 12, 8]). Though a useful simplifying assumption, it forces us to ignore the important facial photo-effects like cast shadows and specularities, and can lead to photo-unrealistic images (Fig. 7). Methods that work without making the limiting Lambertian assumption, still generally ignore cast shadows ([4, 7, 9, 1]) or require a very large number of input images ([6, 9, 1]). Recently a technique that accounts for cast shadows and specularities was proposed by Barmpoutis et al. [2], this technique however can lead to smoothed-out images.

In this paper we propose a novel technique for producing high quality relit images of faces, given a few example (≥ 9) images as input with known illumination directions, which can accurately render cast shadows and specularities. It does not require – manual intervention (unlike [4, 12]), customized data acquisition (unlike [6, 9]) or 3D shape information (unlike [4, 3, 12]). We demonstrate that even in this constrained setting our method can outperform state-of-the-art non-Lambertian ([2]) and Lambertian [3] techniques in terms of image quality (both quantitatively and qualitatively).

2 ABRDF Field Estimation

In order to account for cast shadows and specularities, we work with Apparent Bidirectional Reflectance Function (ABRDF) which describes the variation in pixel intensities as a function of illumination and viewing directions. Here we are concerned with illumination variation hence ABRDF here refers to a spherical function of illumination directions (i.e. fixed pose). Once the ABRDF field of a human face has been captured, novel images of that face can be rendered by sampling its ABRDF field in desired illumination directions. We obtain an accurate estimate of facial ABRDF via a two step process. The first step estimates ABRDFs at each pixel on a subject's face using the input images of the subject via Spline Modulated Spherical Harmonics (SMSH). In the second step we improve the ABRDF representation at each pixel by projecting it in a low variance subspace for ABRDF representation (whose basis functions are called Eigenbubbles) learned using ABRDFs from various other people's faces. Using these two steps we fuse background knowledge about human faces into the ABRDF field estimation process to achieve high quality relit images. This is the first time that a learning based approach has been used in conjunction with facial ABRDFs.

2.1 Spline Modulated Spherical Harmonic

Since ABRDFs are spherical functions, we represent them using spherical harmonic basis, which are a natural orthonormal basis for the spherical domain. We denote the real spherical harmonic basis functions as

Ψ_l^m (order: l, degree: m), with $l = 0, 1, 2, ...$ and $-l \leq m \leq l$:

$$\Psi_l^m(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_{l|m|}(\theta, \phi)\Phi_m(\theta, \phi). \quad (1)$$

where $P_{l|m|}$ are the associated Legendre functions and $\Phi_m(\theta, \phi)$ is defined as

$$\Phi_m(\theta, \phi) = \begin{cases} \sqrt{2}\cos m\phi & m > 0, \\ 1 & m = 0, \\ \sqrt{2}\sin |m|\phi & m < 0. \end{cases} \quad (2)$$

We choose to use only the odd ordered basis function due to the nature of the ABRDF functions and since we are interested in capturing their behavior only on the frontal hemisphere of directions [2].

For a fixed pose, each pixel location has an associated ABRDF and across the whole face, we have a field of such ABRDFs. To model a field of spherical functions ($\mathbb{S}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$), we propose to use modulated spherical harmonics by combining spherical harmonic basis *within* a single pixel and B-splines basis *across* the field. This leads to the following expression for the SMSH

$$\widetilde{\Psi}_l^m(\theta, \phi, \vec{x}, i, j) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} N_{i,4}(x_1) N_{j,4}(x_2) \cdot P_{l|m|}(\theta, \phi)\Phi_m(\theta, \phi), \quad (3)$$

with $\Phi_m(\theta, \phi)$ and $P_{l|m|}$ are as defined before, $\vec{x} = (x_1, x_2)$ are the spline control points, $N_{i,4}$ and $N_{j,4}$ are the cubic B-Spline bases and i and j are the bases indices. We have chosen to use the commonly used bi-cubic B-spline as empirically it provides adequate amount of smoothness for facial ABRDFs.

There are two distinct advantages of using SMSH for ABRDF field estimation. Firstly, the built-in smoothness provides a degree of tolerance to noise which is very common when dealing with image data and secondly, it allows us to use neighborhood information while estimating ABRDF at each pixel location.

2.2 ABRDF Field Estimation using SMSH

Using $S_x(\theta, \phi)$, the given data samples (intensity values) in (θ, ϕ) direction at location $\vec{x}$, the ABRDF field can be estimated by minimizing the following error function:

$$E(w_{ijlm}) = \sum_{\vec{x},\phi,\theta} \|(\sum_{l\in\mathcal{T}} \sum_{m=-l}^{l} \sum_{i,j} w_{ijlm}(\vec{x}) \cdot \widetilde{\Psi}_l^m(\theta, \phi, \vec{x}, i, j)) - S_x(\theta, \phi)\|^2, \quad (4)$$

where the first term in the summation is the representation of the ABRDF function using SMSH. $\mathcal{T}$ is the set of odd natural numbers and w_{ijlm} are the unknown coefficients of the ABRDF field that we seek.

Here, the spline control grid is overlaid on data grid (pixels) and the inner summation on i and j is over the bi-cubic B-Spline basis domain. We minimize this objective function using the non-linear conjugate gradient method (initialized with a unit vector field), for which derivative can be analytically calculated. In this paper we have used 3rd order harmonic approximation which leads to 10 unknown coefficients. When exactly 9 images are used as input an additional constraint that minimizes the norm of the solution is also used. A detailed analysis of the impact of input image distribution is presented in the Sec. 3. We must point out that this is first time that Spherical Harmonics are used in conjunction with B-Splines to represent ABRDF field. In that past spherical Harmonics have only been used with the limiting Lambertian assumption ([3, 10, 12, 8]).

2.3 Eigenbubbles

So far we have not used any domain (facial ABRDF) specific information in our ABRDF representation. To this end, we seek a few salient spherical functions, which define a natural subspace for representing the ABRDFs. We obtain such spherical functions as basis of the low variance subspace derived via spectral analysis of many ABRDFs (from different people and locations). We call them Eigenbubbles. Note that since we have used a spherical harmonics for the ABRDF representation, the distance computation between spherical functions required as part of the spectral analysis, can be carried out using the simple L_2 distance. The use of L_2 distance would not be valid in any other non-orthonormal representation.

Given a bag of ABRDFs represented by their SMSH coefficients $\{\alpha_i\}$, we define the mean ABRDF as $\hat{\alpha} = \sum_i \alpha_i/N$. The data covariance matrix can then be defined as $\mathbf{C} = \sum_i (\alpha_i - \hat{\alpha})(\alpha_i - \hat{\alpha})^T/N$. Next we decompose the square matrix C into its eigenvectors($\mathbf{V}$) and eigenvalues($\mathbf{U}$) as $\mathbf{C} = \mathbf{VUV}^T$. Arranged in the increasing order of corresponding eigenvalues, the eigenvectors - in our case Eigenbubbles - define a low variance subspace for the ABRDF representation.

Global Eigenbubbles are defined to be the ones obtained by putting ABRDFs from all different locations and different individuals into the initial bag of ABRDFs. On the other hand, Local Eigenbubbles are those obtained separately for each pixel, by considering only the ABRDFs lying at the same (roughly registered) pixel locations, from all the given faces. These two definitions would allow us to analyze the impact of the use of Eigenbubble for ABRDF representation in the next section.

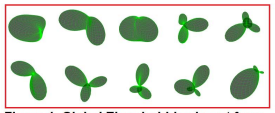

Figure 1. Global Eigenbubbles learnt from Extended Yale B face database. Eigenvalues increase from left to right and then top to bottom.

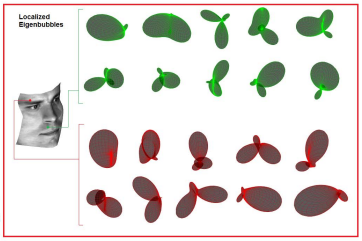

Figure 2. "Local Eigenbubbles" learned from two different locations on a face.

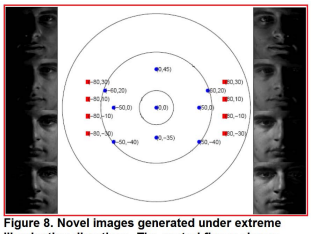

Figure 8. Novel images generated under extreme illumination directions. The central figure shows the illumination directions of the input images as blue circles and the illumination directions of the novel images as red squares.

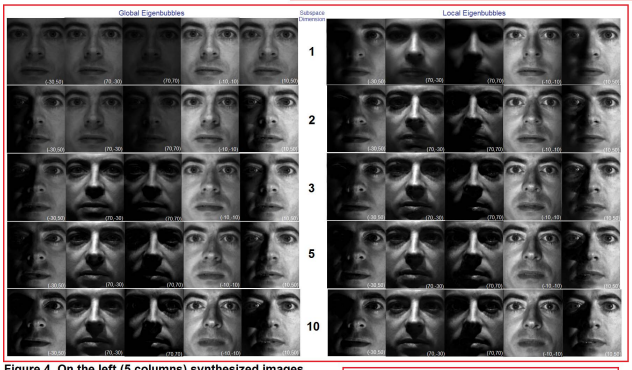

Figure 4. On the left (5 columns) synthesized images using the "Global Eigenbubbles" are shown. From top to bottom, each row uses 1, 2, 3, 5 and 10 Eigenbubbles respectively for its ABRDF field representation. The same setup is repeated on the right using the "Local Eigenbubbles."

Methods	N	Set 1&2	Set 3	Set 4	Total
PAMI 1993 Correlation [2]	4	0.0	23.3	73.6	29.1
CPVR 1994 Eigenfaces [2]	6	0.0	25.8	75.7	30.4
CVPR 2001 Linear subspace [2]	7	0.0	0.0	15.0	4.7
PAMI 2001 Cones-attached [2]	7	0.0	0.0	8.6	2.7
PAMI 2001 Cones-cast [2]	7	0.0	0.0	0.0	0.0
PAMI 2005 9PL [2]	9	0.0	0.0	2.8	0.8
PAMI 2006 3D SH [2]	1	0.0	0.0	2.8	0.8
IJCV 2008 Harmonic (SFS) [2]	1	0.0	0.0	12.8	4.0
CVPR 2008 Tensor Splines [2]	9	0.0	0.0	1.6	0.5
Eigenbubbles	9	0.0	0.0	0.9	0.29

Table 1. Face recognition errors rates. N: Number of input images.

D	Raw	Eigenbubbles	Reduction Ratio	Error/Pixel
1	2016 KB	127.6 KB	6.25 %	18.3
2	2016 KB	253.6 KB	12.50 %	15.0
3	2016 KB	379.6 KB	18.75 %	13.8
4	2016 KB	505.6 KB	25.00 %	13.9
5	2016 KB	631.6 KB	31.24 %	13.8
6	2016 KB	757.6 KB	37.51 %	14.2
7	2016 KB	883.6 KB	43.76 %	13.3
8	2016 KB	1009.6 KB	50.01 %	13.2
9	2016 KB	1135.6 KB	56.26 %	13.1
10	2016 KB	1261.6 KB	62.51 %	12.9

Table 2. ABRDF field compression. D: Subspace dimension. Errors are intensity errors per pixel.

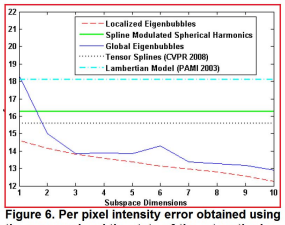

Figure 6. Per pixel intensity error obtained using the proposed and the state-of-the-art methods.

Figure 5. The first four rows show images of subjects from CMU PIE while rest show images of subjects from Extended Yale B. The (azimuth,elevation) angles are at the bottom right corner of each image. For CMU PIE input directions were (32,2), (0,-9), (32,-9),(0,3),(-32,-8),(-32,3),(38,6),(0,10),(-32,5).

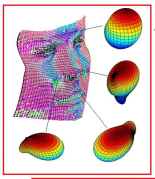

Figure 3. Estimated ABRDF functions at various locations on a human face.

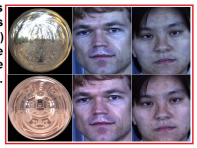

Figure 9. First column shows the light probes (Eucalyptus Grove and St. Peter's Basilica) used to light the images of the 2 CMU PIE subjects in the subsequent two columns.

Figure 7. Qualitative comparison between the proposed technique, the Anti-symmetric tensor splines model [2] and the Lambertian model [3].

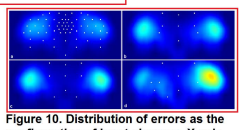

Figure 10. Distribution of errors as the configuration of input changes. X-axis represents azimuth and Y-axis represents elevation angles. "Hotter" colors show larger errors. The white squares represent the directions of illumination in images used as input.

Notes:
1. Large images embedded, please zoom to see details.
2. Figure numbers are jumbled to fit images better.

3 Experiments & Discussion

Relighting: We used the Extended Yale B [8] and the CMU PIE[11] benchmark databases for our experiments. Foremost we show the shape of learned Global and Local Eigenbubble in Fig. 1 and Fig. 2 respectively. It can be noted that the Local Eigenbubbles vary from one location to another depending on the local nature (shadowing etc.) of the ABRDFs. In Fig. 3 we present the actual ABRDFs estimated at a few locations on the face represented using Global Eigenbubbles. It can be seen that the ABRDFs have more complicated shapes than what the Lambertian half-cosine can represent.

To analyze the representational power of the Global and the Local Eigenbubbles, in Fig. 4 we present samples of the ABRDF field estimated using Eigenbubbles (9 image input, shown in Fig. 8). The images have been synthesized using novel illumination directions whose azimuthal and elevation angles are mentioned in the bottom right corner of each image. From these images, foremost, it can be noted that as the subspace dimension is increased so does the visual quality of the images. Secondly, quality of the images produced using the Local Eigenbubbles, especially for low dimensional subspaces, is better than when the Global Eigenbubbles are used. Thirdly, for both types of Eigenbubbles, visually high quality images are rendered for 5 and higher dimensional representations. Accurate depiction of shadows, both attached and cast, and specularities can be readily noted from the images.

Next we test our representation across different faces from both the databases in Fig. 5. Our framework can be readily extended to color images by finding Eigenbubbles for each color channel separately. Using faces belonging to different races, we have demonstrated the

capability of the Global Eigenbubbles to represent the ABRDFs of surfaces which can have somewhat different surface properties. Note that the shadows have been crisply generated and the specularities have been meaningfully rendered. In Fig. 6 we quantitatively examine Eigenbubbles against the Lambertian [3] and the Antisymmetric Tensor Splines [2]. For this experiment 9 images were used to estimate ABRDF fields for subjects from Extended Yale B dataset using various methods. Samples of these fields in the 64 directions for which ground-truth data is available were used to compute per pixel intensity errors. It can be noted that the Lambertian model, with its limited representative power in the presence of cast shadows and specularities performs the worst. As expected, the Local Eigenbubble outperforms the Global Eigenbubbles, but both of these techniques, outperform the state-of-the-art Antisymmetric Tensor Spline method. The improvement in the image quality with increased number of subspace dimensions can also be noted. The above trend can be qualitatively noted in Fig. 7 where shadows and specularities are more crisply rendered in Eigenbubbles than the state-of-the-art methods. We present results for extreme lighting conditions in Fig. 8. The central figure shows the illumination directions of the input images (**same for all Yale results**) as blue circles and the illumination directions of the novel images as red squares. The 8 images corresponding to the red squares are shown to the left and the right.

Our method can also be used to render images under complex lighting conditions. In Fig. 9 we present images of 2 subjects rendered using Eucalyptus Grove and St. Peter's Basilica light probes [5] (shown in the first column). 9 images for each subject were used as input and we used 2500 samples of the light probes to generate these images. Finally we examine the impact of the distribution of input images on the final synthesized image quality in Fig. 10. The cases (a) to (d) plot 4 different distribution of input images (white dots). In each image, X-axis represents azimuth and Y-axis represents elevation angles. Images rendered in 64 directions in all the cases were used to compute per pixel intensity errors which are plotted with hotter color representing larger errors. In case of uniformly distributed input ((a)-(c)), larger errors are primarily towards the extreme angles (image boundaries) while in the case of biased input distribution (d), bias in errors can also be noted.

Face Recognition: In a recognition setup, if 9 gallery images are available, our method can be used to render new images from estimated ABRDF field to augment the gallery set. We present recognition error rates obtained using such a scheme in Table 1. It can be noted that Eigenbubbles improve upon the errors reported in [2] when used with simple Nearest Neighbor classifier.

ABRDF Field Compression: A novel application of our method is the compression of the ABRDF fields. The central thesis is that if we can achieve good image quality with (lets say) 5 dimensional Global Eigenbubble subspace, instead of storing all 10 or more SMSH coefficients, we can store just 5. For 64 images of Extended Yale B dataset, it take 2 MB of memory to store the raw 192×168 images, while the space savings in various subspace representations are presented in Table 2. It must be noted that our technique cannot be compared with image compression methods as ABRDF fields can generate as many images as desired while image compression techniques cannot.

4 Conclusions

In this paper we have presented a novel mathematical framework for spherical function field representation using SMSH and used it in conjunction with novel Eigenbubbles to accurately represent facial ABRDF fields. We have shown that our method out performs state-of-the-art ([2]) both in terms of image quality, per pixel errors and face recognition rates. We have also presented a novel application of our framework to ABRDF field compression.

References

[1] N. Alldrin and T. Zickler. Photometric stereo with non-parametric & spatially-varying reflectance. *CVPR*, 08.

[2] 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*, 2008.

[3] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, 2003.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *PAMI*, 25(9):1063–1074, 2003.

[5] P. Debevec. Rendering synthetic objectsinto real scenes. *SIGGRAPH*, 1998.

[6] P. Debevec, T. Hawkins, C. Tchou, H. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. *SIGGRAPH*, 2000.

[7] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying brdfs from photometric stereo. *ICCV*, 2005.

[8] K. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *PAMI*, 27(5):684–698, 2005.

[9] J. Lee et al. A bilinear illumination model for robust face recognition. *ICCV*, 2005.

[10] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. *SIGGRAPH*, 01.

[11] 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, RI, 2001.

[12] L. Zhang, S. Wang, and D. Samaras. Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable model. *CVPR*, 2005.