FACE REPRESENTATION UNDER DIFFERENT ILLUMINATION CONDITIONS

1Haitao Wang, 2Hong Wei, 1Yangsheng Wang

1Institute of Automation, Chinese Academy of Sciences
2Dept. of Computer Science, The University of Reading
Email: htwang@nlpr.ia.ac.cn; h.wei@reading.ac.uk; wys@nlpr.ia.ac.cn

ABSTRACT
To deal with image variations due to illumination problem, recently Ramamoorthi and Basri have independently derived a spherical harmonic analysis for the Lambertian reflectance and linear subspace. Their theoretical work provided a new approach for face representation, however both of them had the assumption that the 3D surface normal and albedo are known. This assumption limits this algorithm’s application. In this paper, we present a novel method for modeling 3D face shape and albedo from only three images with unknown light directions and this work well fills the blank, which Ramamoorthi and Basri left. By taking the advantage of similar 3D shape of all human faces, the highlight of the new method is that it circumambulates the linear ambiguity by 3D alignment. The experiment results show that our estimated model can be perfectly employed to face recognition and 3D reconstruction.

1. INTRODUCTION

Illumination problem is one of the most challenging tasks in computer vision. For face images, gray level variation due to the illumination changes may significantly exceed the differences due to different identities. Furthermore slight changes in lighting conditions will cause distinct variations in face image gray value distribution. This is the reason why illumination is listed in the table of the main difficulties for face recognition [1].

To cope with this problem, traditional (Principle Component Analysis) PCA-based [3] approach needs densely sampled images under all illumination conditions. This is computationally intractable and impractical. Therefore models built by PCA algorithm are always biased for a scarcity of sample images. In addition, this approach can be only applied to point light sources. This is contrary to the real world, where light sources are usually area lights and have largely continuous lighting distributions.

To represent face under all possible lighting conditions, many researchers [4][5][6][10] have done a large number of experiments. Their empirical study showed that human face, as a typical example of Lambertian object, does lie in a low-dimensional subspace and the first five or seven principle components can explain the image variations due to lighting changes very well. However there is no theoretical foundation to support their explanations.

Recently, two researchers, Ramamoorthi [7][8] and Basri [9] have independently worked out the mechanism of image variations under different illumination conditions by spherical harmonic theory. By their spherical harmonic analysis results, we do not need sampling all the images produced by varied lighting conditions, such as in the case of illumination cone [12] and the light source can be any type, e.g. point, area, or any combination of them. Though their work has provided a new method for face representation, both of them assume the known 3D surface normals and albedo (or unit albedo) or use complex 3D Scanner. This limits this algorithm’s application. In this paper, we present a novel method for modeling face’s 3D surface normals and albedo from only three images and apply the face models into 3D reconstruction and recognition. The arrangement of this paper is as follows. In section 2, we briefly describe the spherical harmonic representation of face images and elaborate our modeling process. In section 3, the surface normal estimated in previous section is utilized for 3D reconstruction and the nine harmonic images are employed for recognition. In section 4, we make a conclusion.

2. FACE REPRESENTATION BY SPHERICAL HARMONICS AND NEW MODELING PROCESS

Experiments by Hallinan [10], and Epstein [4] have demonstrated that human face images under different lighting conditions lie in a low-dimension subspace, which can be approximated by PCA. But their works were based on empirical data and there was not theoretical foundation. Until recently Ramamoorthi [7][8] and Basri [9] independently worked out a theoretical analyzed
relationship between the low-dimensional subspace and the spherical harmonic images, which indicates that the complex illumination problem in computer vision can be explained in theory.

2.1. Spherical Harmonic Representation

In this subsection, we briefly introduce the work of [7][8][9]. Let’s assume that human face \( F \) is a convex Lambertian object and is illuminated by distant isotropic light sources. Let \((x, y, z)\) and \((\theta, \phi)\) be the local coordinate and the spherical coordinate of point \( P \) on the surface of \( F \) respectively. The relationship between them is

\[
(x, y, z) = (\cos \theta \sin \phi, \sin \theta \cos \phi, \cos \phi)
\]

where \( 0 \leq \theta \leq \pi \), \( 0 \leq \phi \leq 2\pi \).

The reflectance of point \( p \) by a point light source \( I(\theta, \phi) \) in the half sphere is defined by

\[
r(\theta, \phi) = \int_0^{2\pi} \int_0^\pi k(\theta)L(\theta, \phi) \sin \theta d\theta d\phi
\]

where \( k(\theta) = \max(\cos \theta, 0) \).

And we can refer equation (2) as a convolution

\[
r(\theta, \phi) = k * l
\]

The spherical harmonics are a set of functions that form an orthonormal basis on the sphere, which is analogue to Fourier basis convoluting in the plane. These functions are denote by \( h_{nm} \), with \( n = 0, 1, 2, \ldots \), and \( -n < m \leq n \).

\[
h_{nm}(\theta, \phi) = N_{nm} P_{nm}(\cos \theta) \exp(\text{Im}\phi)
\]

where \( N_{nm} \) is a normalization factor and \( P_{nm} \) are the Legendre functions.

The reflectance function can be rewritten as

\[
K(\theta, \phi) = \sum_{n,m} K_{nm} H_{nm}(\theta, \phi)
\]

\[
l(\theta, \phi) = \sum_{n,m} l_{nm} H_{nm}(\theta, \phi)
\]

\[
r(\theta, \phi) = \sum_{n,m} C_{nm} H_{nm}(\theta, \phi)
\]

Analytical study of Ramamoorthi [7][8] and Basri [9] shows that \( k_n \) vanishes for odd values of \( n \geq 1 \) and the even terms fall to zero rapidly. Moreover the second order approximation (first nine terms) captures more than 99% of the energy. For more detail of \( k_{nm}, h_{nm} \) and \( r_{nm} \), one can reference [7][8][9].

Consequently the spherical harmonic bases for reflectance function are the first nine harmonics.

\[
h_{00} = 0.2821;
\]

\[
h_{11}, h_{10}, h_{1-1} = 0.4866(n_x, n_y, n_z);
\]

\[
h_{20} = 0.3154(3n_z^2 - 1);
\]

\[
h_{22} = 0.5462(n_x^2 - n_y^2)
\]

where \((n_x, n_y, n_z)\) is the surface normal of face.

Therefore the harmonic images, which form a base for represent face images under all light conditions, are constructed as follows:

\[
b_{nm}(p) = \rho h_{nm}(n_x, n_y, n_z)
\]

where \( \rho \) is the face albedo.

2.2. Our New Method of Modeling

According to the of the Lambertian Model [3], a face image with attached shadow \((n^T s < 0)\) can be described as

\[
I(x, y) = \max(0, \rho, (x, y)n(x, y)^T s)
\]

where \( I \) is the \( p \times 1 \) face image with all the \( p \) pixels concatenated by row or by column, \( \rho \) \((> 0)\) is the \( p \times 1 \) albedo(surface texture), \( n^T \) is the \( p \times 3 \) surface normal of face and \( s \) is the \( 3 \times 1 \) light source vector with unit intensity.

We assume that we have three images for each face with equal intensity, non-collinear light source direction and small shadow. Under this assumption, we can model the face albedo and surface normal with linear ambiguity.

Let \( I_1, I_2, I_3 \) are the three images for a face, \( n^T \) is the surface normal for this face and \( s_1, s_2, s_3 \) are the corresponding light source direction of the three images. Using the singular value decomposition (SVD) [6][11], we get

\[
J = [I_1, I_2, I_3] = \rho n^T [s_1, s_2, s_3] = SVD(J)
\]

\[
= UDV = \rho n^T S^\star
\]

where \( D \) is a diagonal matrix whose elements are the square roots of the eigenvalues of \( JJ^T \). \( U \) is the \( p \times 3 \) matrix, the columns of which are corresponding to the normalized eigenvectors of the matrix \( JJ^T \). And \( V \) is the \( 3 \times 3 \) matrix, whose columns are the eigenvectors of \( JJ^T \).

Since the matrix \( J \) is composed of three image, we can transform the UDV form into \( \rho n^T S^\star \) form, where

\[
S^\star = V, \quad B = U, \quad \rho = ||B|| \quad \text{and} \quad n^T = \begin{bmatrix} b_1/||B|| \\ \cdots \\ b_p/||B|| \end{bmatrix}
\]

Because \( V \) is orthogonal and has unit length and the three light sources are not collinear, the column \( V \) can be seen as the three rotated light sources with unit length, the rank of which must be 3.

There is an invertible linear ambiguity matrix \( A \), which satisfies
\[ n^T = n_i^T A \]  \hspace{1cm} (11)
and
\[ S = A^{-1} S^* \]  \hspace{1cm} (12)
where \( n^T \) and \( S \) are the true surface normal and light direction of face image \( I \).

This linear ambiguity is belonging to a subset of GBR [14]. According to Belhumeur [14], this ambiguity does not affect the images formed by \( n^T \) and \( S^* \). This result is obvious because the ambiguity matrix \( A \) rotates \( n^T \) and \( S \) at the same time, therefore their dot product does not change. In fact, this ambiguity can be discard in the face model, such as in the formation of illumination cone in [9]. However, if this ambiguity is not solved, we can not calculate the accurate harmonic images for the face model.

Let assume \( n_0^T \) is a face’s true surface normals and all the face have similar 3D shape. If each face surface normals \( n_i^T \) with is aligned according to the \( n_0^T \), then we can get an estimation of the true surface normals.

Therefore a simple way to circumambulate this problem is to align all face models by \( C_i \) under the assumption that all faces have similar 3D shape. Let \( E \) be the alignment energy function,
\[ E = (n_0^T - n_i^T C_i)(n_0^T - n_i^T C_i)^T \quad i = 1, \ldots, m \]  \hspace{1cm} (13)
where \( m \) is the number of face model, \( n_0^T \) is the alignment model and \( n_i^T \) the face model with ambiguity.

Minimizing this function by least squares technique, we get
\[ C_i = (n_i^T n_i^T)^{-1} n_i^T n_0^T . \]  \hspace{1cm} (14)

Then all the surface normals are aligned according to \( n_0^T \) by
\[ n_i^T = n_i^T C_i \]  \hspace{1cm} (15).

Our algorithm is similar to that of [15], but they do not exclude the albedo effects.

3. EXPERIMENTS AND DISCUSSION

Two kinds of experiments, face recognition and 3D reconstruction are done by face model estimated by 3D alignment.

3.1 Recognition

Our recognition method is very straightforward by checking the correlation coefficient between the input image \( I \) and the synthetic images, which are from face models in the database and are synthesized according to the input image’s illumination condition.

For face recognition, because our aim is to find ways for dealing with face image variation due only to light direction changes, we assume that all the face images is frontal and there is no other variation, except light direction variation. And thereby all the images used are manually segmented and aligned.

The Yale B database used contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions. Because our focus is illumination problem, only frontal pose is considered and all the 640 images are manually segmented, rotated and aligned to avoid other factors, except illumination, to affect the images. And then we sort these 640 images into 4 sets according to the illumination direction, which is expressed by azimuth and elevation with respect to the camera axis.

To verify the effectiveness of our algorithm, we use the same classifier for spherical harmonic model estimated by our algorithm (nine harmonic images and four harmonic images), direct correlation method and quotient image [16] method. The classifier used here is the nearest neighbor classifier, which employs correlation coefficient as the measurement for classification. The recognition results shown in Table 1. It is clear that our estimated model (nine harmonic images) has almost perfect recognition rate, which is comparable to that of the complex illumination cone [12]. Even in the four harmonic images case, this model still demonstrates its robustness for lighting changes in Set 3 and Set 4. The decreasing of recognition rate from Set 1 to Set 4 indicates that the representation effects deteriorate with the increasing azimuth or elevation. This result tells us that neglecting the cast shadow makes the spherical representation only valid when the azimuth and elevation is not too large.

Our recognition results are similar to that of Kuang-Chih Lee [18], but we have a more simple modeling process.

<table>
<thead>
<tr>
<th>Table 1. Recognition Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Nine harmonic images</td>
</tr>
<tr>
<td>Four harmonic images</td>
</tr>
<tr>
<td>Direction Correlation</td>
</tr>
<tr>
<td>Quotient image</td>
</tr>
</tbody>
</table>

3.2 Reconstruction

Figure 1 shows one example of 3D reconstruction from our estimated surface normals. For reconstruction, the face model must be aligned by true surface normals. The
reconstruction result (Figure 1) by our estimated surface normals is excellent compared with most SFS [17] algorithms.

Though Basri [13] has also described a more general algorithm for solving linear ambiguity and applied it into 3D reconstruction, our method is more practical and simpler for face images reconstruction and has similar effects.

Figure 1. the 3D surface reconstruction (a) from the unknown light source images (b) according to surface normals (c) estimated by 3D alignment

4. CONCLUSION AND FURTHER WORK

In this paper, we introduce a new simple and practical approach for face modeling. By taking advantage the similarity of human faces, this method makes detour of the linear ambiguity problem by means of 3D alignment. The recognition and reconstruction experiments indicate the effectiveness of this algorithm.

Though the spherical harmonic function can clearly derive a theoretical basis for face images under varied lighting, it ignores the cast shadow, which has great effects on face images in cases of large light direction variation. Therefore in the further, we wish to take the cast shadow into account by the 3D surface reconstructed though graphic techniques.

Acknowledgement: The Conference Participation is Support by Nokia Bridge the World Program.

REFERENCES