

Illumination Normalization for Face Recognition Using Energy Minimization Framework

Xiaoguang T\textsuperscript{(a)}, Nonmember, Feng YANG\textsuperscript{††}, Member, Mei XIE\textsuperscript{†††}, and Zheng MA\textsuperscript{†}, Nonmembers

SUMMARY Numerous methods have been developed to handle lighting variations in the preprocessing step of face recognition. However, most of them only use the high-frequency information (edges, lines, corner, etc.) for recognition, as pixels lied in these areas have higher local variance values, and thus insensitive to illumination variations. In this case, information of low-frequency may be discarded and some of the features which are helpful for recognition may be ignored. In this paper, we present a new and efficient method for illumination normalization using an energy minimization framework. The proposed method aims to remove the illumination field of the observed face images while simultaneously preserving the intrinsic facial features. The normalized face image and illumination field could be achieved by a reciprocal iteration scheme. Experiments on CMU-PIE and the Extended Yale B databases show that the proposed method can preserve a very good visual quality even on the images illuminated with deep shadow and high brightness regions, and obtain promising illumination normalization results for better face recognition performance.

key words: energy minimization, illumination normalization, face recognition

1. Introduction

Recently, numerous methods, e.g., DWT [1], logarithmic total variation (LTV) model [2], WeberFace [3] etc., have been proposed to handle lighting variations in the preprocessing step of face recognition. However, many of them keep focus on the high-frequency components (e.g., lines, edges) of the observed face images. Namely, the high-frequency subbands are used for recognition while the low-frequency contents are discarded. This may cause the loss of some useful detailed information in the normalized images. Furthermore, without these intrinsic facial details, it’s certainly hard to generate a frontal-illumination image with a good visual quality, and hence weaken the effectiveness of many image processing methods and steps.

Other kinds of methods are also proposed to address the illumination normalization, which aims at restoring the face images to a condition of normal illumination with a compensating scheme, e.g., Histogram Equalization (HE) [4], Gamma correction [5], etc. However, most of these methods do illumination normalization directly on the whole image and rarely consider the illumination condition of the processed image, which is difficult to obtain reasonable results with good visual quality. In [6], Xie X. et al. proposed a new illumination normalization method that merges the components from both low and high frequency. It was shown that their normalized images contain more facial details and preserve good visual quality. However, their work may not be able to properly tackle the images with extreme lighting condition, e.g., deep shadow and high brightness regions. In the later work, Wang H et al. [7] proposed to use robust Principal Component Analysis to remove the shadow of the high-frequency features based on Xie X.’s work. Nevertheless, this algorithm still has the same problem of poorly handling extreme lighting condition.

In this paper, we propose a new method for illumination normalization by using an energy minimization framework. Our method tries to remove the illumination field of the observed face images while preserve the intrinsic facial details. As a result, the proposed method could preserve very good visual quality. The normalized results and the illumination field could be achieved by minimizing the energy function during the process of reciprocal iteration. After applying our algorithm to images under various illumination conditions, we use the normalized images for face recognition on both CMU-PIE and Extended Yale B face databases, and shows satisfactory performance.

2. Energy Minimization Based Illumination Normalization Method

2.1 Motivation and Decomposition of Images

A face image \( I \) can be decomposed as the product of two multiplicative components: the true image (image without illumination effects) and the received light (illumination field)[8]. Taking the additive noise into consideration, this can be modeled as

\[
I(x) = J(x)L(x) + n(x)
\]

where \( I(x) \) is the intensity of the observed face image at pixel \( x \), \( L(x) \) is the illumination field, \( n(x) \) is additive noise with zero-mean, and \( J(x) \) is the true image which can be taken as the normalized one. The nature of \( L(x) \) is determined by the illumination source, whereas \( J(x) \) is determined by the characteristics of the image. Therefore, the normalized face image (true image) could be obtained by
estimating the illumination field \( L \) and then dividing the image \( I \) by it. However, it is impossible to estimate \( L \) from \( I \), unless something else is known about either \( L \) or \( J \). A common assumption is that the illumination varies smoothly in the illumination field, while the edges keep the same between the original image and the normalized one [9]. Under this assumption, the true image \( J \) can be estimated as the ratio map of the observed image and the smoothly varying property could be used for the estimation of the illumination field \( L \).

Theoretically, a function could be approximately by a linear combination of a set of basis functions up to arbitrary accuracy. In this case, we represent \( L \) by a linear combination of a given set of smooth basis functions \( G(x) = (g_1(x), \ldots, g_M(x))^T \), which ensures the smoothly varying property of the illumination field [10]. Therefore, the estimation of illumination field \( L \) could be performed by finding the optimal coefficients \( w_1, \ldots, w_M \) in the linear combination \( L(x) = \sum_{k=1}^{M} w_k g_k \) by a column vector \( w = (w_1, \ldots, w_M)^T \), as follows.

\[
L(x) = w^T G(x) \tag{2}
\]

For the estimation of the true image \( J \). We take \( J \) as the association of multiple fuzzy membership functions [11]. Each membership function can be regarded as a percentage map with the same size as the observed image. We denote \( u_i \) the \( i \)-th membership function, where \( u_i \) take values either 0 or 1 and satisfy \( \sum_{i=1}^{N} u_i(x) = 1 \), \( N \) is the number of the memberships, and \( c_i \) the weight assigned to \( u_i \). In this case, the \( J \) can be approximated by

\[
J(x) = \sum_{i=1}^{N} c_i u_i(x) \tag{3}
\]

If the appropriate membership functions \( u = (u_1, \ldots, u_N)^T \), the optimal coefficient vector \( w = (w_1, \ldots, w_N)^T \) and weight vector \( c = (c_1, \ldots, c_N)^T \) are found, the true image \( J \) will be estimated accordingly.

2.2 Energy Formulation of Illumination Normalization

In view of model (1), We consider the problem of finding the multiplicative components \( J \) and \( L \) of an observed image \( I \), and using \( L_2 \) for cost calculation, such that the following energy is minimized

\[
E(J, L) = \int_{\Omega} |J(x) - L(x)J(x)|^2 dx \tag{4}
\]

where \( \Omega \) is the image domain. However, if no constraints are imposed on variables \( J \) and \( L \), it will be an ill-posed problem to minimize this energy. To make the problem solvable, the search spaces of \( J \) and \( L \) should be confined by exploiting some knowledge about the two unknown components. In fact, the properties of the illumination \( L \) and the true image \( J \) have already been discussed above and can be exploited to confine the search space of \( L \) and \( J \). With the representations in Eq. (2) and Eq. (3), the energy function \( E(J, L) \) can be expressed in terms of three variables [11], \( c = (c_1, \ldots, c_N)^T \), \( u = (u_1, \ldots, u_N)^T \), and \( w = (w_1, \ldots, w_M)^T \):

\[
E(c, u, w) = \int_{\Omega} |J(x) - w^T G(x) \sum_{i=1}^{N} c_i u_i(x)|^2 dx \tag{5}
\]

The optimization of \( L \) and \( J \) can be achieved by minimizing the energy \( E \) with respect to \( c, u, \) and \( w \).

Recall of the constraints of \( u, \sum_{i=1}^{N} u_i(x) = 1, \) and \( u_i \) take values either 0 or 1. When \( N \) is big enough, the Energy \( E \) can be approximated as

\[
E(c, u, w) \approx \sum_{i=1}^{N} \int_{\Omega} |J(x) - w^T G(x)c_i|^2 u_i(x) dx \tag{6}
\]

By exchanging the order of integration and summation, this can be described as

\[
E(c, u, w) = 2 \int_{\Omega} \sum_{i=1}^{N} d_i(x) u_i(x) dx \quad \sum_{i=1}^{N} d_i(x) u_i(x) = \bar{J}(x) - J(x) \tag{7}
\]

where \( d_i(x) = \frac{1}{2} |J(x) - w^T G(x)c_i|^2 \) is the least squares (LS) cost function. We introduced the idea of adding a regularization term to LS to confine the search spaces. So that the new energy takes the form

\[
E(c, u, w) = \int_{\Omega} \sum_{i=1}^{N} (|J(x) - w^T G(x)c_i|^2 + \mu ||w||^2) u_i(x) dx \tag{8}
\]

where \( \mu \) is a very small regularization coefficient that controls the relative importance of the cost function and the regularization term.

2.3 Energy Minimization

The energy minimization can be achieved by alternately minimizing \( E(u, w, c) \) with respect to each of its variables given the other two remain fixed. For fixed \( u \) and \( w \), there is a unique minimizer of \( E \) with respect to \( c \), denoted by \( \hat{c} \). It’s easy to obtain the solution drive form the Euler-Lagrange equation of energy \( E(u, c, w) \) [10]

\[
\hat{c}_i = \frac{\int_{\Omega} J(x) L(x) u_i(x) dx}{\int_{\Omega} L^2(x) u_i(x) dx}, \quad \text{for} \quad i = 1, \ldots, N \tag{9}
\]

Given \( u \) and \( c \), there is a unique minimizer of energy \( E \) with respect to \( w \), denoted by \( \hat{w} \). It can be shown that

\[
\hat{w} = \frac{\int_{\Omega} G(x) I(x) J^{(1)}(x) dx}{\int_{\Omega} (G(x) I(x) + \lambda I) J^{(2)}(x) dx} \tag{10}
\]

where \( J^{(1)}(x) = \sum_{i=1}^{N} c_i u_i(x), J^{(2)}(x) = \sum_{i=1}^{N} c_i^2 u_i(x) \).

Given the variables \( w \) and \( c \) fixed, the energy \( E(u, c, w) \)
is minimized when \( \mathbf{u} = \hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_N) \) as below:
\[
\hat{u}_i(x) = \begin{cases} 
1 & \text{if } i = i_{\min}(x); \\
0 & \text{if } i \neq i_{\min}(x);
\end{cases}
\]
(11)

where
\[ i_{\min}(x) = \arg \min_i \delta_i(I(x)), \quad i = 1, \ldots, N. \]
(12)

and \( \delta_i(I(x)) = |I(x) - w^T G(x)c_i|^2 + \mu ||w||^2 \).

The proposed method is iteratively conducted, and the whole process can be described by the following pseudocode:

**Algorithm 1** The proposed algorithm to solve Eq. (8)

1. **Initialization:** Initialize the membership functions \( u_i \) by random matrices where each entry follows a uniform distribution in [0, 1].
   Set \( c_0 = 0, w_0 = 1 \).
2. **repeat:**
3. \( \delta \) by Eq. (9);
4. Update the membership functions \( u_i \) by Eq. (11) and Eq. (12);
5. Update \( w \) by Eq. (10).
6. until the termination criterion: \( ||\delta - \Delta\delta|| \leq \epsilon \) is reached.


\( \delta \) denotes the Euclidean distance, and \( \epsilon \) is a small positive number defined by the user.

### 3. Experimental Results and Analysis

In this section, we present the simulation results of the proposed algorithm on illumination normalization of face images, and then use the normalized images for face recognition. CMU-PIE [12] and Extended Yale B [13] face databases are used for evaluation. Images in CMU-PIE are obtained from 68 subjects and captured with, and without background lighting, each subject with 21 images. 1428 (21 × 68) frontal face images are selected for testing in our experiments. The Extended Yale B database contains 2432 (64 × 38) frontal face images with 38 individuals. Each individual is taken under 64 illumination conditions from 9 views. In our experiments, only the frontal face images from these two databases are selected. All images are simply aligned and each image is resized to 100 × 100.

#### 3.1 Comparisons of the Visual Quality

We first show the illumination normalization results of the proposed method in Fig. 1. As can be seen, the normalized images keep a good visual quality. The bright regions are effectively corrected and the deep shadow are almost completely removed, indicating that the proposed method could reduce illumination effect while simultaneously preserving the intrinsic facial properties.

In Fig. 2, we compare the visual results of normalized images between the proposed method and four baseline algorithms, i.e., histogram equalization (HE) [4], logarithmic total variation (LTV) [2], and R&S (LOG-DCT & NPL-QI) [6]. HE aims at transforming the distribution of the pixel intensity values of the input image into a uniform distribution and consequently at improving the images global contrast. LTV only make use of the small-scale facial features of face images (e.g., lines, edges) while discarding the large-scale components. R&S (LOG-DCT & NPL-QI) are algorithms which combined large- and small- scale features by employing LOG-DCT and NPL-QI, respectively. As can be observed, the proposed method could preserve more facial details in the normalized images and achieve better normalization performance on visual quality. Although R&S (NPL-QI) can keep the facial details unchanged to some extent, it handles poorly in deep shadow regions and some of the bright regions, which may weaken the effectiveness of many image processing methods and steps.

#### 3.2 Quantitative Evaluation Results

Two evaluation metrics, peak signal to noise ratio (PSNR) and structural similarity (SSIM) index are used to quantitatively evaluate the normalization performance between the proposed method and the baseline methods. PSNR has been proved to be a good estimation of the perceived quality. A larger PSNR value provides a higher image quality, and a small PSNR value implies high numerical differences between images. SSIM is a well-know quality metric used to measure the similarity between two images, and is considered to be correlated with the quality perception of the human visual system. The closer the SSIM value is to 1, the better the visual quality is. In our experiments, The image which is under normal illumination condition is taken as the ground truth to compute PSNR and SSIM. For Ex-
tended Yale B, the face image of any subject captured by the frontal camera is the ground truth. For CMU-PIE database, we choose 42 (21*2) images for shown, 1st to 21st are the 21 images of different illumination conditions with background light off, while 22nd to 42nd are the same images of conditions with background light on. For Extended Yale B database, all the 64 images of different illumination condition of each individual are involved in our experiments. The quantitative results are reported in Fig.3. As shown, the PSNR and SSIM curves obtained by our method are higher than those achieved by the other methods, suggesting our method achieves the normalization results with better visual quality.

3.3 Classification with Face Images

In this subsection, recognition results are reported for evaluation of the proposed Method. To make the recognition task challenging, only one face image of each individual, the one under normal illumination condition, is registered as the reference image. The remaining images regarding the same individual are used for testing. To avoid the impact of any feature extraction/selection schemes, the simplest classifier, Nearest Neighbor (NN) classifier is selected for face classification. We also give the result of original images without any processing (ORI) for comparison. The recognition results on CMU-PIE, Extended Yale B and “CMU-PIE + Extended Yale B” are tabulated in Table 1. As can be seen, the proposed method improve the overall average recognition rate on CMU-PIE, Extended Yale B, and ‘CMU-PIE + Extended Yale B’ from 34.4% to 83.4%, 42.9% to 86.4%, and 36.8% to 82.2%, respectively, significantly better than HE, LTV, and R&S (LOG-DCT), and as good as R&S (NPL-QI).

4. Conclusion

In this paper, we have proposed a novel energy minimization framework with the purpose of illumination normalization for face recognition. The normalized face image and illumination field could be simultaneously achieved during the process of reciprocal iteration. The advantage of the proposed method is that it can effectively remove the illumination field while keeping very good visual quality. Experiments on CMU-PIE and Extended Yale B databases have illustrated the effectiveness and advantages of the proposed method on the normalization under varying illumination. The face images normalized by the proposed method have also been used for face recognition and achieved better performance compared with several other baseline algorithms.

Table 1 Comparisons between different methods for face recognition on CMU-PIE and Extended Yale B databases

<table>
<thead>
<tr>
<th>Method</th>
<th>Face Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CMU-PIE</td>
</tr>
<tr>
<td>ORI</td>
<td>34.4%</td>
</tr>
<tr>
<td>HE</td>
<td>46.6%</td>
</tr>
<tr>
<td>LTV</td>
<td>78.1%</td>
</tr>
<tr>
<td>RLS &amp; (LOG-DCT)</td>
<td>81.3%</td>
</tr>
<tr>
<td>RLS &amp; (NPL-QI)</td>
<td>82.0%</td>
</tr>
<tr>
<td>Our Method</td>
<td>83.4%</td>
</tr>
</tbody>
</table>

References