A New Method of Illumination Normalization for Robust Face Recognition

Young Kyung Park, Bu Cheon Min, and Joong Kyu Kim

School of Information and Communication Engineering, SungKyunKwan University.
300, Chun-Chun-Dong, Chang-An-Ku, Suwon, Korea 440-746
{multipym, jkkim}@skku.edu

Abstract. In this paper, we propose a novel method of illumination normalization developed on the basis of the retinex theory. In retinex based methods, illumination is generally estimated and normalized by first smoothing the input image and then dividing the estimate into the original input image. The proposed method estimates illumination by iteratively convolving the input image with a 3×3 averaging mask weighted by an efficient measure of the illumination discontinuity at each pixel. In this way, we could achieve a fast illumination normalization in which even face images with strong shadows are normalized efficiently. The proposed method has been evaluated based on the Yale face database B and the CMU PIE database by using PCA. Carrying out various scenarios of test, we have found that our method presented consistent and promising results even when we used images with the worst case of illumination as training sets. We believe that the proposed method has a great potential to be applied to real time face recognition systems, especially under harsh illumination conditions.

1 Introduction

Among many factors affecting the performance of face recognition systems, illumination is known to be the one of the most significant. It has already been shown that illumination causes larger variations in intensity of face images than pose. For example, ambient lighting varies greatly during the course of the day, and from one day to another, as well as between indoor and outdoor environments. Moreover, strong shadows cast from a direct light source can make certain facial features invisible. Therefore, illumination normalization is a major requirement in the face recognition process and also is a central topic in the field of computer vision. In recent years, numerous approaches have been proposed to solve illumination problems in face recognition. Belhumeur et al. [1] extended the eigenface algorithm of Turk and Pentland [2] to fisherfaces by employing a classifier based on Fisher's linear discriminant analysis. They reported that fisherfaces outperform eigenfaces under conditions of varying illumination. Georghiades et al. [3] showed that the illumination cones of human faces can be approximated well by low dimensional linear subspaces. Therefore, the set of face images in fixed pose but under different illumination conditions can be efficiently represented using an illumination cone. However, the above methods either require

certain assumptions to be made about the light source or need a large number of training sets, and these requirements are not considered practical in real applications. On the other hand, there are alternative methods available which are based on retinex theory. The retinex theory motivated by Land [4] is based on the physical imaging model, in which an image $I(x, y)$ is regarded as the product $I(x, y) = R(x, y)L(x, y)$ where $R(x, y)$ is the reflectance and $L(x, y)$ is the illumination at each pixel (x, y). Here, the nature of $L(x, y)$ is determined by the illumination source, whereas $R(x, y)$ is determined by the characteristics of the imaged objects. Therefore, the illumination normalization for face recognition can be achieved by estimating the illumination 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 R. Hence, various assumptions and simplifications about L, or R, or both are proposed to solve this problem [5]. A common assumption is that edges in the scene are edges in the reflectance, while illumination spatially changes slowly in the scene. Thus, in most retinex based algorithms, the reflectance R is estimated as the ratio of the image I and its smooth version which serves as the estimate of the illumination L, and many smoothing filters to estimate the illumination have been proposed for robust face recognition. Single Scale Retinex (SSR), the latest version of Land's retinex that was implemented and tested by Jobson et al. [6], employed a simple linear filter with Gaussian kernel. However, halo effects are often visible at large illumination discontinuities in I, and Gross and Brajovie [5] introduced an anisotropic filter to reduce these halo effects to some extent. More recently, self-quotient image (SQI) [7] has been proposed with impressive improvement of performance for illumination problem. SQI employs the weighed Gaussian filter in which the convolution region is divided into two sub-regions with respect to a threshold, and separate values of weights are applied in each sub-region. These retinex based methods have common advantages that they do not require training images and has relatively low computational complexity. However, these methods cannot completely remove large illumination discontinuities, and more fast method is required for real time face recognition system. In this paper, we propose a novel method for illumination normalization based on the retinex theory. The key idea of our method is to estimate illumination by iteratively convolving an input image with a 3×3 averaging mask weighted by an efficient measure of the illumination discontinuity at each pixel. The paper is organized as follows. In section 2, the proposed algorithm is described in detail. In section 3, the experimental results are presented. Finally a conclusion is made in section 4.

2   Illumination Normalization

Since our method is based on the retinex theory, the process of our method for a pixel (x, y) in an image is described as

$$R(x, y) = \frac{I(x, y)}{L(x, y)} \qquad (1)$$

As mentioned in section 1, the illumination L is estimated as a smooth version of input image I. Jobson et al. [6] evaluated various functional forms for the smoothing filter and found that the Gaussian form performed better than the inverse square suggested by Land [4]. For face recognition, however, Gaussian filter is known not to satisfy a robustness requirement, which smoothing must be carried out among pixels which having homogeneous illumination. This robustness requirement implies that the estimated illumination must be discontinuous at locations where the input image I has strong discontinuities of intensity [8]. This robust requirement is to more accurately take real world scenes into account, where illumination discontinuities such as shadows frequently occur. On the other hand, the Gaussian kernel, which is currently in wide use for smoothing, has an inherent computational limitation to be applied in real time; that is the computational complexity increases proportionally to the square of the kernel size, causing the overall speed of the smoothing filter very slow. To ensure a fast smoothing satisfying the robustness requirement at the same time, we present a novel method based on iterative convolution.

2.1   Iterative Convolution

Iterative convolution is a fast algorithm, in which an efficient short-length convolution is performed iteratively to build a long convolution [9]. As shown in table 1, iterative convolution is very efficient in terms of the computational complexity. Moreover, it was shown that, by virtue of the central limit theorem, convolving

Table 1. Computational complexity of two convolution methods

	Iterative short-length	Fixed long-length
Complexity	$O(nN^2k^2)$	$O(N^2K^2)$
	N: image size	N: image size
	k: small kernel size($2\sim3$)	K: kernel size
	n: # of iteration	

a 3×3 averaging kernel n times with an image I approximates the convolution of I with Gaussian kernel of $\sigma = \sqrt{n/3}$ and size of $3(n+1) - n = 2n + 3$ [10]. We can formulate the iterative convolution process as follows. Let $I^{(0)}(x, y)$ be the input image at each point (x, y) before smoothing. Then, the smoothed image $I^{(t+1)}(x, y)$ at the $(t+1)th$ iteration is given by: [9]

$$I^{(t+1)}(x, y) = \frac{1}{N^{(t)}} \sum_{i=-1}^{1} \sum_{j=-1}^{1} I^{(t)}(x + i, y + j)w^{(t)}(x + i, y + j) \qquad (2)$$

with

$$N^{(t)} = \sum_{i=-1}^{1} \sum_{j=-1}^{1} w^{(t)}(x + i, y + j) \qquad (3)$$

where

$$w^{(t)}(x, y) = 1, \ \forall(x, y) \text{ and } \forall t. \tag{4}$$

$N^{(t)}$ in (3) represents a normalizing factor, and $w^{(t)}(x, y)$ in (4) indicate the corresponding weights of the convolution mask. This filter smoothes the image everywhere, even across discontinuities because this is just an approximated version of Gaussian filter. Consequently, halo effects might often be visible when we apply the Gaussian filter. Therefore, although iterative convolution is efficient in computational complexity, it doesn't meet the robustness requirement. For robust smoothing, the $w^{(t)}(x, y)$, thus, must be set adaptively based on discontinuities of intensity values. We will present the adaptive weighting method in the next subsection
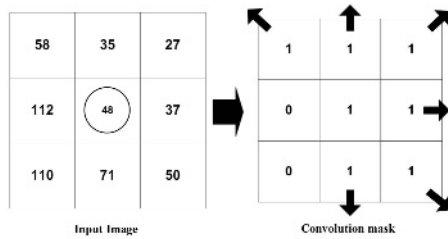

Fig. 1. An example of a convolution mask constructed by applying adaptive weighting ($\alpha = 0.2$) to an input image

2.2   Adaptive Weighting

To set weights of convolution mask $w^{(t)}(x, y)$ adaptively, we must know where illumination discontinuities take place. If we already knew the locations of these discontinuity points, then we could set the corresponding weights of the convolution mask $w^{(t)}(x, y)$ to zero [9]. In the framework of illumination normalization, this means that smoothing must be mainly carried out among pixels with homogeneous illumination. Unfortunately, it is impossible to correctly acquire the locations of illumination discontinuities when only one image is given, and additional information about the scene is not available. To solve this problem, we can set $w^{(t)}(x, y)$ based on the variance between a central pixel and neighboring pixels in convolution mask. Let pixel location (m, n) correspond to a central pixel of a convolution mask with 3×3 size. Then, the variance between the central pixel and neighboring pixels in convolution mask is

$$\tau(m, n) = \sum_{(x,y)\in\Omega} \sum |I(x, y) - I(m, n)| \tag{5}$$

where Ω is a convolution region, and (x, y) indicates the locations of the neighboring pixels. The corresponding weights of the convolution mask $w^{(t)}(x, y)$ are determined according to the variance as follows:
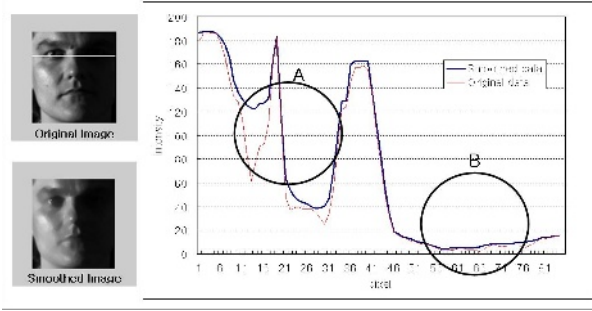
Fig. 2. The intensity profiles of horizontal lines in an original image and a smoothed image after 9 iterations

$$w^{(t)}(x,y) = \begin{cases} 1 & |I(x,y) - I(m,n)| < \alpha\tau(i,j) \\ 0 & \text{else} \end{cases} \tag{6}$$

where α determines the level of strong discontinuity. If α approaches 0, level of strong discontinuity which must be preserved during the process of smoothing decreases. As shown in figure 1, only pixels with relatively little different intensity values from a current pixel are convolved. Once weights of the convolution mask corresponding to each pixel of an input image are determined, iterative convolution is performed. Due to the diffusion characteristic of the iterative convolution, there are two significant operations affecting the image as the iteration proceeds: one is the preservation of strong discontinuities, and the other is the smoothing within regions with small variance. Figure 1 shows these operations well, where diffusion effects of smoothing occur only in the directions of arrow. Now, for more accurate description of real environments, we address an additional constraint that surfaces cannot reflect more light than what is shed on them [8]. Thus, the reflectance R must be always smaller than one. By this constraint, (2) can be modified as follows.

$$I^{(t+1)}(x,y) = \frac{1}{N^{(t)}} \sum_{i=-1}^{1} \sum_{j=-1}^{1} I^{(t)}(x+i, y+j)w^{(t)}(x+i, y+j)$$

$$I^{(t+1)}(x,y) = \max\{I^{(t+1)}(x,y), I^{(t)}(x,y)\} \tag{7}$$

Figure 2 shows a result of smoothing using the proposed iterative convolution with adaptive weighting with $\alpha = 0.2$. Each region A and B in the graph correspond to two eyes in the face. We can see that regions of facial features are efficiently smoothed while the illumination discontinuities are preserved. Note that the signal variation under light region is much larger than that under shadow region, and the facial features (in this case, the two eyes) have far different scales in these two regions. For a good and stable performance of face recognition, therefore, it is necessary that the small scale signal in the shadow region must be amplified so that both the light and shadow regions have similar signal scales
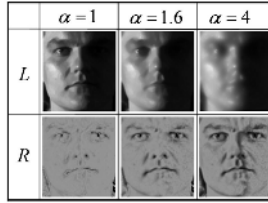
Fig. 3. Effect of different α in adaptive weighting. The first row shows estimated illumination using the proposed smoothing method, and the second row shows final results which are obtained by dividing an input images by estimated illuminations.

after illumination normalization. For this purpose, we apply the retinex theory formulated in (1), and figure 3 shows the results according to some typical values of α. In order to ensure a robust face recognition, the smoothing must be carried out at larger scale than the scale of facial features; that is, we should emphasize the small scale facial features such as mouth, nose, eyes, and eyebrows while removing the illumination. And this mainly depends on the proper choice of α; for example, if α is too small, facial features would not be visible. By some investigations as shown in figure 3, we selected a proper value of $\alpha = 0.2$ for our experiments which will be described in the following section. The number of iterations is also important factor for robust face recognition. If the number of iterations is too large, normalization effects would decrease with the increase of visual effects. For our experiment, we fixed the number of iterations to $n = 9$.

3 Experimental Results and Discussion

In order to evaluate the robustness and effectiveness of the proposed method, we used the images from two publicly available databases: Yale face database B [3] and CMU PIE database [11], and computed the recognition accuracies using Eigenfaces (Principal Component Analysis (PCA)) [2]. The proposed method is also compared with three other existing methods of illumination normalization: SSR [6], SQI [7] and histogram equalization. We present the respective test result for each database.

3.1 Yale Face Database B

The Yale face database B contains 5,760 images taken from 10 subjects under 576 viewing conditions(9 poses $\times$ 64 illumination conditions). Since we are only concerned with the illumination problem in this paper, we selected 640 images for 10 subjects representing 64 illumination conditions under the frontal pose. Images in the database are divided into 5 subsets based on the angle of the light source directions. The 5 subsets are subset 1($0°$ to $12°$), subset 2($13°$ to $25°$), subset 3($26°$ to $50°$), subset 4($51°$ to $77°$), and subset 5(above $78°$) [3]. We refer to images in both subset 4 and subset 5 as images with large illumination

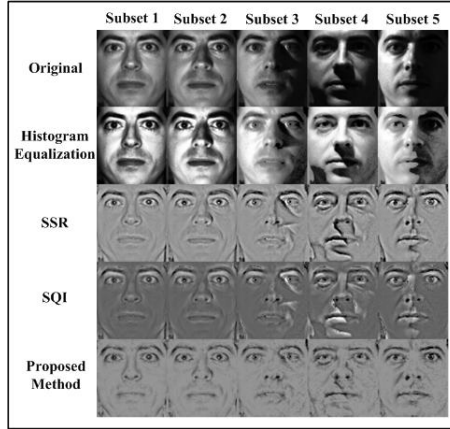**Fig. 4.** Representative images of 10 subjects in Yale face database B

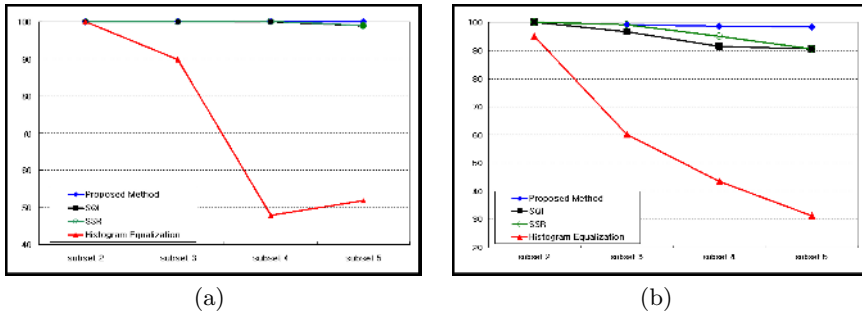**Fig. 5.** Illumination normalization effect comparison

Fig. 6. Recognition accuracies(%) on the Yale face database B using (a)subset 1, and (b)only the representative images, as the training set

variation. Among the 640 images, we discarded 7 corrupted images and finally constructed total 633 images (70, 118, 118, 138, 189 images in subset 1 to 5). Figure 4 shows the representative images of the 10 subjects, and figure 5 shows illumination normalization results by different methods according to each subset of the one among 10 subjects. To apply our proposed method to a face recognition system, we first used subset 1 as training set and tested other subsets. Figure 6(a) shows the recognition results using each illumination normalization method.

Table 2. Average recognition rate (%) comparison on the Yale face database B using images in all different subsets (from subset 2 to subset 5) as the training set

	Proposed method	SQI	SSR	Histogram equalization
Recognition rate	97.65%	93.11%	94.06%	42.36%

(a)

(b)

(c)

(d)
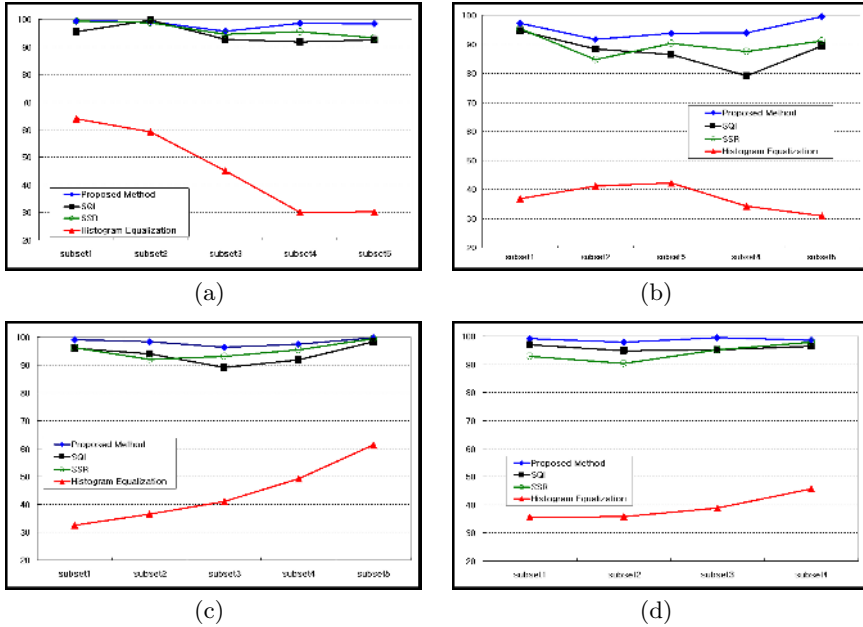
Fig. 7. Recognition accuracies(%) on the Yale face database B using (a)subset 2 ($test2$), (b)subset 3 ($test3$), (c)subset 4 ($test4$), and (d)subset 5 ($test5$) as the training set

The proposed method achieved recognition rates of 100% in all subsets. As a second test, we only used the 10 images(one for each subject in subset 1) in figure 4 as a training set and tested other subsets. The results are given in figure 6(b), and it is clear that the proposed method outperforms other methods. Finally, we used images in each of the other subsets (2 to 5) as training set and computed the recognition accuracy [12]. Since subset 2 to 5 represent the illumination conditions close to real environments, we can say that this test is more meaningful and practical than the other two previous tests. As shown in figure 7, four tests are carried out. Each $testX$ denotes that the training set is from subset X, where $X = 2, 3, 4, 5$, and the test is done for the remaining subsets. Table 2 compares recognition rates by averaging all results from these four tests. From the figure 7 and the table 2, it is clear that the proposed method has consistent and promising results even when images with large illumination variation (subset 4 and subset 5) are used as the training set.
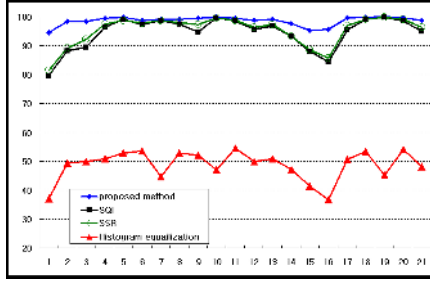
Fig. 8. Recognition accuracies(%) on the CMU PIE database

Table 3. Average recognition rate (%) comparison on the CMU PIE database

	Proposed method	SQI	SSR	Histogram equalization
Recognition rate	98.57%		94.53%	95.32%	48.69%

3.2   CMU PIE Database

The CMU PIE database contains 41,368 images obtained from 68 subjects. We took images of frontal faces with 21 different illumination conditions. Among 68 subjects, we removed one subject because it was not a frontal image. Thus, the total number of images we used for our test is 1,407.

The recognition rates were computed by the k-fold strategy: i.e, k images of each subject (in our test, $k = 1$) are selected for training and the remaining $21 - k$ images of each subject are selected for test. This process was repeated $(21 - k + 1)$ times by changing training images. Figure 8 shows recognition rates according to each trial. The proposed method also outperforms other methods and has consistently high recognition rates over 95%.

4   Conclusion

In this paper, we proposed a new illumination normalization method for robust face recognition. The proposed method, which is based on the retinex theory, estimates the illumination using a novel iterative smoothing methodology, and turned out to be superior to existing smoothing methods in terms of both time complexity and performance. The essence of our proposed method is that the smoothing is carried out by iteratively convolving an input image with a 3×3 averaging mask weighted by a efficient measure of the illumination discontinuity at each pixel. Using the proposed method, we showed that even images with strong shadow are effectively normalized, and consequently the recognition accuracies notably outperformed the existing methods. Especially, our proposed method presented consistent and promising results even when we used images

with large illumination variations as training set. We believe that the proposed method has a wide range of applications in real time face recognition systems, especially under harsh illumination conditions.

References

1. Belhumeur,P.N., Hespanha,J., Kriegman,D.J.: Eigenfaces vs. Fisher-faces:Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence,20(7), (1997)
2. Turk, M., Pentland, A.: Eigenfaces for Recognition.Journal of Cognitive Neuroscience, 3(1):71-86,(1991)
3. Georghiades, A. S., Belhumeur, P. N., Kriegman, D. J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6):643-660, (2001)
4. Land, E.: An alternative technique for the computation of the designator in the retinex theory of color vision. Proc. Nat. Acad. Sci., vol. 83, 3078-3080, (1986)
5. Gross, R., Brajovic, V.: An image preprocessing Algorithm for illumination invariant face recognition. In 4th International Conference on Audio and Video Based Biometric Person Authentication, (2003)
6. Jobson, D. J., Rahman, Z., Woodell, G. A.: Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing, vol. 6, 451-462, (1997)
7. Wang, H., Li, S. J., Wang, Y.: Generalized quotient image. IEEE CVPR, (2004)
8. Shaked, D., Keshet, R.: Robust recursive envelope operators for fast retinex. Hewlett-Packard Research Laboratories Technical Report, HPL-2002-74R1, (2002)
9. Saint-Marc, P., Chen, J.-S., Medioni, G.: Adaptive smoothing: a general tool for early vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(6):514-529, (1991)
10. Trucco, E., Verri, A.: Introductory techniques for 3-D computer vision, Prentice Hall, (1998)
11. Sim, T., Baker, S., Bsat, M.: The CMU Pose,Illumination,and Expression Database.IEEE Trans on PAMI, vol. 25, 1615-1618, (2003)
12. Chen, T., Yin, W., Zhou, X. S., Comaniciu, D., Huang, T. S.: Illumination Normalization for Face Recognition and Uneven Background Correction Using Total Variation Based Image Models. IEEE CVPR, (2005)