Face Recognition under Varying Lighting Conditions Using Self Quotient Image

[1]Haitao Wang, [2]Stan Z Li, [1]Yangsheng Wang

[1]Institute of Automation, Chinese Academy of Sciences, Beijing, 100080, China,
Email: {htwang,wys}@nlpr.ia.ac.cn

[2]Beijing Sigma Center, Microsoft Research Asia
Beijing 100080, China
Email: szli@microsoft.com

Abstract

In this paper, we introduce the concept of Self-Quotient Image (SQI) for robust face recognition under varying lighting conditions. It is based on the Quotient Image method [4][5] to achieve lighting invariant. However, the SQI has three advantages: (1) It needs only one face image for extraction of intrinsic lighting invariant property of a face while removing extrinsic factor corresponding to the lighting. (2)No alignment is needed. (3) It works in shadow regions. The theoretical analysis on conditions where the algorithm is applicable and a non-iterative filtering algorithm for computing SQI are presented. Experiment results demonstrate the effectiveness of our method for robust face recognition under varying lighting conditions.

1. Introduction

Lighting variation is one of the most difficult problems for face recognition and has received much attention [1-13] in recent years. It is well known that image variation due to lighting changes is more significant than that due to different personal identities [12]. Recent years, many algorithms have been proposed; let us highlight three major approaches of lighting modeling for 3D object recognition: as Illumination Cone [1]-[3], Quotient Image [4][5], and Spherical Harmonic Subspace [6-10].

The Illumination Cone method theoretically explained that face images due to varying lighting directions form an illumination cone. In this algorithm, both self-shadow and cast-shadow were considered and its experiment results outperformed most existing methods.

Ramamoorthi [6-8] and Basri [9] [10] independently developed the spherical harmonic representation. This original representation explained why images of an object under different lighting conditions can be described by low dimensional subspace in some previous empirical experiments [16-18]. Given enough training images with same pose and different lightings for 3D modeling, the above two approaches achieved almost perfect recognition rate on some face databases [3][19]. However, the requirement of large training sets restricts their applications in many areas.

Whereas the above two approaches assumed that each individual had a different 3D geometry and needed a training set for each person, quotient image proposed by Shashua and Riklin-Raviv [4][5] is a simple yet practical algorithm for extracting lighting invariant representation. It was shown that the quotient image, i.e. image ratio between a test image and linear combination of three non-coplanar illuminated images, depended only on the albedo information, which is illumination free.

Still another approach for dealing with lighting problem is from the 2D image processing viewpoint. Jobson, et al [13][14] presented a multi-scale version of Retinex method [20] for high quality visual display of high dynamic image on low dynamic devices, such as printer and computer screen. This is closely related to the lighting issue. More recently Gross and Brajovie [15] presented an anisotropic version of Retinex for lighting normalization. Both the above two groups of authors proposed an algorithm which estimates low frequency component of the input image as the light field and compensated illumination variations by subtracting it from the input image.

We would like to differentiate between extrinsic and intrinsic factors in imaging process and analyze their influence on object recognition. The Lambertian model states that the image of a 3D object is subject to three factors, namely the surface normal, albedo and lighting. These can be factorized into an intrinsic part of the surface normal and albedo, and an extrinsic part of lighting. The identity of an object is determined by the intrinsic factor only and this in fact largely motivated the existing work [1-11] on lighting normalization. While pure appearance based learning methods, such as PCA, ICA and LDA, learn appearance models from examples mixing intrinsic and extrinsic factors, our analysis explains why they cannot achieve high performance with features extracted from the learned models.

In this paper, we analyze of the Retinex algorithm in terms of Lambertien model and demonstrate its relationship with the quotient image. Based on this, we propose a novel concept called the self-quotient image (**SQI**) which is defined as the ratio of the input image and its smooth versions. We analyze properties of the **SQI**, and show that **SQI** is a lighting invariant representation of 3D objects. Compared with the quotient image, the **SQI** has the following advantages: (1) no need for alignment; (2) valid for both shadow and non-shadow region; (3) valid for any type of lighting sources.

The paper is organized as follows: Section 2 analyzes the related approaches, Retinex method and the quotient image method. Section 3 establishes the concept of **SQI**, presents our multi-scale anisotropic approach for **SQI** and its relationship with **QI**, and analyzes its properties in different imaging cases. The experiment results are analyzed in Section 4 and finally the conclusion is made in section 5.

2. Related Approaches

Retinex algorithms are based on the reflectance-illumination model (1) instead of Lambertian model (4).

$$I = RL \tag{1}$$

where $\boldsymbol{I}$ is the image, $\boldsymbol{R}$ is the reflectance of the scene and $\boldsymbol{L}$ is lighting.

The lighting $\boldsymbol{L}$ can be considered as the low frequency component of image $\boldsymbol{I}$, as has been theoretically proved by the spherical harmonics analysis [6-10], and hence can be an estimation by using a low-pass filter

$$L \approx F * I \tag{2}$$

where $\boldsymbol{F}$ is a Gaussian filter and * donates the convolution operation. The general form of the center/surround Retinex can be defined as

$$R = \frac{I}{L} = \frac{I}{F * I} \tag{3}$$

The Retinex in [13][14] was designed for dynamic range compression, which was applied in displaying high dynamic range image onto some low dynamic range device as printer and screen. They processed images according to acquire the best visual effect and there is no quantitative standard to measure Retinex's results.

The quotient image method [4][5] is designed for dealing with lighting changes in face recognition. It provides an invariant representation of face images under different lighting conditions. The theory is based on the Lambertian model

$$I(x, y) = \rho(x, y) n(x, y)^T s \tag{4}$$

where $\boldsymbol{\rho}$ is the albedo (surface texture) of face, $\boldsymbol{n(x,y)^T}$ is the surface normal (3D shape) of the object (same for all objects of the class), and $\boldsymbol{s}$ is the point light source, which may vary arbitrarily.

The quotient image $\boldsymbol{Q_y}$ of face $\boldsymbol{y}$ against face $\boldsymbol{a}$ is defined by

$$\begin{aligned} Q_y(u,v) &= \frac{\rho_y(u,v)}{\rho_a(u,v)} = \frac{\rho_y(u,v) n(u,v)^T s_y}{\rho_a(u,v) n(u,v)^T s_y} \\ &= \frac{I_y}{\rho_a(u,v) n(u,v)^T \sum_j x_j s_j} \\ &= \frac{I_y}{\sum_{j=1}^{3} I_j x_j} \end{aligned} \tag{5}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ range over the image, $\boldsymbol{I_y}$ is an image of object $\boldsymbol{y}$ with the illumination direction $\boldsymbol{s_y}$, and $\boldsymbol{x_j}$ are combining coefficients estimated by Least Square based on training set [4][5], and $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, $\boldsymbol{I_3}$ are three non-collinearly illuminated images.

The following assumptions are made in the quotient image framework: (a) the imaging process follows the Lambertian model *without* shadow and the object is illuminated by a point light source; (b) all the faces under consideration have the same shape, i.e. the same surface normal; (c) there is no shadow in face images; (d) accurate alignment between faces is known; and (e) a training set of faces under at least three non-collinear lighting are available as basis for estimation of lighting directions.

However, in a face recognition system the above assumptions could not be satisfied at the same time.

For example, the light sources are generally not of point; 3D face shapes of different people are not the same in general; the shadow can exist; and accurate alignment is still an unsolved problem by now.

In the following, we will introduce what we call the *self-quotient image*, and demonstrate that it has lighting invariant properties similar to those of the original quotient image, yet presents several technical advantages.

3. SQI and its Lighting Invariant Property Analysis

3.1 Definition of SQI and its Invariant Property Analysis

In the following, we define the self-quotient image as an intrinsic property of face images of a person.

Definition 1: (Self-Quotient Image **SQI**) The Self-Quotient image $\boldsymbol{Q}$ of image $\boldsymbol{I}$ is defined by

$$Q = \frac{I}{\hat{I}} = \frac{I}{F * I} \tag{6}$$

where $\hat{I}$ is the smoothed version of $\boldsymbol{I}$, $\boldsymbol{F}$ is the smoothing kernel, and the division is point-wise as in the original quotient image. We call $\boldsymbol{Q}$ the Self-Quotient Image because it is derived from one image and has the same quotient form as that in the quotient image method. We will demonstrate in the following part of this section that SQI has similar lighting invariant property as that of quotient image. But there are significant differences between quotient image and self-quotient image. (1) Sself-quotient image is calculated from one image, (2) only image processing technique is used in self-quotient image and no empirical learning is needed, and (3) there is no assumption about face images

The lighting invariant properties are demonstrated below using the Lambertian model but with shadow. When shadows present, the Lambertian model with shadows can be represented as

$$I = \min(\rho n^T s, 0) \tag{7}$$

We consider three cases of different shapes and shadow conditions in the analysis of SQI.

Case 1: In regions without shadow and with small $\boldsymbol{n^T}$ variations. In this case, $n^T(u,v)s \approx C_1$, where $\boldsymbol{C_1}$ is a constant. Then we have

$$Q = \frac{I(u,v)}{\hat{I}(u,v)} \approx \frac{\rho(u,v)C_1}{[\rho(u,v) * F]C_1} = \frac{\rho(u,v)}{\rho(u,v) * F} \tag{8}$$

In this case, $\boldsymbol{Q}$ is approximately illumination free and depends only on the albedo of the face. Equation (8) is similar in form to the quotient image; however it is calculated only from the self image.

Case 2: **In regions without shadow but with large $\boldsymbol{n^T}$ variation.** In this case, $n^T(u,v)s$ is not a constant. The **SQI** is

$$Q = \frac{I(u,v)}{\hat{I}(u,v)} = \frac{\rho(u,v)n^T(u,v)s}{F * [\rho(u,v)n^T(u,v)s]} \tag{9}$$

In such regions, $\boldsymbol{Q}$ depends on the 3D shape, albedo and lighting $\boldsymbol{n^T}$. Therefore $\boldsymbol{Q}$ is not illumination free in this case.

Case 3: In shadowed regions. In these regions, the gray value is low and less variable. We assume that in shadow regions, light is uniformly distributed from all directions, i.e. for any $\boldsymbol{n(u,v)^T}$ in shadow**,** all the visible lights form a semi-hemisphere. Therefore, the summation of the dot products between $\boldsymbol{n^T}$ and $\boldsymbol{s_i}$ is constant in such regions.

$$n(u,v)^T \sum_{i=1}^{\infty} s(u,v)_i = \sum_{i=1}^{\infty} n(u,v)^T s(u,v)_i = C_2 \tag{10}$$

where $\boldsymbol{C_2}$ is a constant. Therefore, $\boldsymbol{Q}$ in shadow regions can be written as equation (8).

As in case 1, **SQI** in this kind of regions is also illumination-free; in other words, the SQI can remove the shadow effect, as shown in Fig. 1.

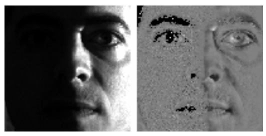


Fig. 1 De-shadow effects of SQI

Although the analysis is based on the Lambertian model with point lighting, it is also valid for other types of lighting sources. This is because any lighting can be expressed as a linear combination of $\boldsymbol{L}$ point lighting sources, as follows

$$I = \rho n^T S = \rho n^T \sum_{i=1}^{L} s_i \tag{11}$$

If we replace the point lighting source $\boldsymbol{s}$ in cases 1 - 3 with $\boldsymbol{S}$ with as the above, the analytic results still hold.

The above analysis shows the following two properties of the self-quotient image: (1) The algorithm is robust to lighting variation for case 1 and 3. (2) SQI is not the expected reflectance as in Retinex, but the albedo ratio in case 1 and case 3 and lighting dependent image ratio in case 2.

For face recognition, if we can ensure that the filter's kernel size is small enough compared with face surface normal $\boldsymbol{n}^{T}$'s variation, the self-quotient image will be illumination free as analyzed previously. However, when the filter's kernel size is too small, SQI will approach one and the albedo information is lost.

Figure 2 shows results of the self-quotient image filtering. The self-quotient image of boy's hair is not as dark as it in the original image; the whole face looks flatter; and shadows are removed.

Figure 2. (left) Original Image (right) Self-Quotient Image

The advantages of the self-quotient method as opposed to the original quotient image is summarized as follows: (1) The alignment between image $\boldsymbol{I}$ and its smoothed version $\hat{I}$ is automatically perfect, and hence it does not need an alignment procedure. (2) No training images are needed for the estimation of the lighting direction because the lighting fields of $\boldsymbol{I}$ and $\hat{I}$ are similar. (3) the self-quotient image is good at removing shadows; whereas in the previous approaches [1-11], the shadow problem was either ignored or was solved by complex 3D rendering. (4) Lighting sources can be any type.

Note that the property of $\boldsymbol{Q}$ is dependant on the kernel size. If the kernel size of $\boldsymbol{F}$ is too small, $\boldsymbol{Q}$ will approximate to one and albedo information will be severely reduced. If the kernel size of $\boldsymbol{F}$ is too large, there will appear halo effects near step-edge region. We use the multi-scale technique to make the result more robust, and in practice, we choose kernel sizes to take more care smoother regions.

3.2 Implementation of SQI

The only processing needed for **SQI** is smoothing filtering. We design a weighed Gaussian filter for anisotropic smoothing, as illustrated in Fig.3, where $\boldsymbol{W}$ is the weight and $\boldsymbol{G}$ is the Gaussian kernel, and $\boldsymbol{N}$ is the normalization factor for which

$$\frac{1}{N}\sum_{\Omega} WG = 1 \tag{12}$$

where $\boldsymbol{\Omega}$ is the convolution kernel size. We divide the convolution region into two sub-regions $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$ with respect to a threshold $\boldsymbol{\tau}$. Assuming that there are more pixels in $\boldsymbol{M_1}$ than in $\boldsymbol{M_2}$, $\boldsymbol{\tau}$ is calculated by

$$\tau = Mean(I_{\Omega}) \tag{13}$$

For the two sub-regions, $\boldsymbol{W}$ has corresponding value.

$$W(i,j) = \begin{cases} 0 & \mathrm{I(i,j)} \in \mathrm{M_2} \\ 1 & \mathrm{I(i,j)} \in \mathrm{M_1} \end{cases} \tag{14}$$

If the convolution image region is smooth , i.e. little gray value variation (non-edge region), there is also little difference between the smoothing the whole region and part of the region. If there is large gray value variation in convolution region, i.e. edge region, the threshold can divide the convolution region into two parts $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$ along the edge and the filter kernel will convolute only with the large part $\boldsymbol{M_1}$, which contains more pixels. Therefore the halo effects can be significantly reduced by the weighted Gaussian kernel.

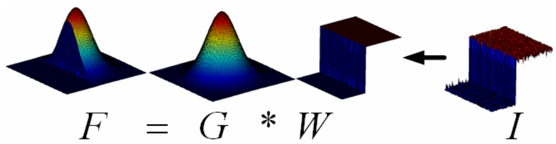


Figure. 3 Anisotropic Smoothing Filter

The essence of this anisotropic filter is that it smoothes only the main part of convolution region i.e. only one side of edge region in case of step edge region.

The division operation in the **SQI** may magnify high frequent noise especially in low signal noise ratio regions, such as in shadows. To reduce noise in $\boldsymbol{Q}$, we use a nonlinear transformation function to transform $\boldsymbol{Q}$ into $\boldsymbol{D}$,

$$D = T(Q) \tag{15}$$

where $\boldsymbol{T}$ is a nonlinear transform.

Jobson, et al [13][14] used Logarithm, which is considered as the similar characteristic of human visual ability. We find in our experiments that Arctangent and Sigmoid nonlinear function have similar or superior results in dynamic range compression for recognition effects.

Our implementation of **SQI** approach is summarized below:

(1) Select several smoothing kernel $\boldsymbol{G_1, G_2, \ldots, G_n}$ and calculate corresponding weights $\boldsymbol{W_1, W_2, \ldots, W_n}$ according to image $\boldsymbol{I}$, and then smooth $\boldsymbol{I}$ by each weighed anisotropic filter $\boldsymbol{WG_i}$.

$$\hat{I}_k = I \oplus \frac{1}{N} WG_k, \; \boldsymbol{k = 1, 2, \ldots, n} \tag{16}$$

Calculate self-quotient image between each input image $\boldsymbol{I}$ and its smoothing version

$$Q_k = \frac{I}{\hat{I}_k}, \; \boldsymbol{k = 1, 2, \ldots, n} \tag{17}$$

(2) Transfer self-quotient image with nonlinear function

$$D_k = T(Q_k), \; \boldsymbol{k = 1, 2, \ldots, n} \tag{18}$$

(3) Summarize nonlinear transfered results

$$Q = \sum_{k=1}^{n} m_k D_k, \; \boldsymbol{k = 1, 2, \ldots, n} \tag{19}$$

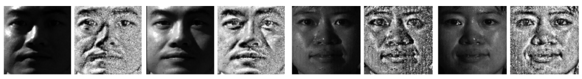
The $\boldsymbol{m_1, m_2, m_n}$ are the weights for each scale of filter and we set them to one in our experiments.

4. Experiments and Discussion

Experiments are performed to evaluate SQI for face recognition, using Yale face database B [3] and CMU PIE face database [19]. Frontal face images with lighting variation are selected from the two face databases to reduce the image changes only due to lighting variations. There are 68 subjects in CMU PIE and we select the frontal face images which were taken under 20 different illuminations without background lighting for each subject. There are 640 images (10 subjects with 64 images each) from Yale B. The eyes, nose and mouth are located manually for each image, and the face is then aligned and cropped. The PCA and original QI methods are also included as the baselines, in which the PCA (60 dimensional) is learned by using all the examples from either PIE or Yale B data sets.

Figure 4 show some results of SQI based lighting normalization. We can see that the convolution based anisotropic filtering is very effective in smoothing the noisy image without blurring the step edge and shadows are removed.

For the PIE data set, the leave-one-out scheme is used, i.e. each image as template in turn and the others as test images. The results are compared in Figure 5 for the 20 different leave-one-out partitions. For the Yale B data set, the images are divided into 4 subsets of increasing illumination angles, and only the frontal illuminated images are used as the templates. The results are shown in Figure 6 for the 4 different data sets. Compared with PCA and original QI, the new algorithms, SQI, can significantly improve the recognition rate both in CMU PIE and Yale B face database.



(a) Yale B (b) CMU PIE

Figure 4. Example Results of **SQI** Light Removal

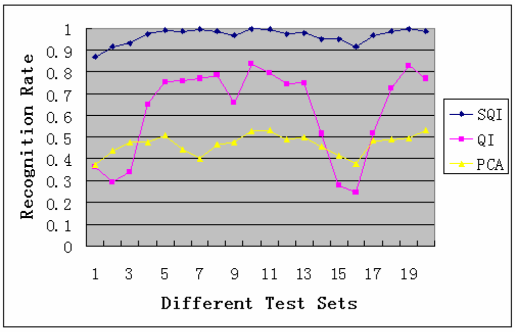


Figure 5. Recognition Results on **CMU PIE**

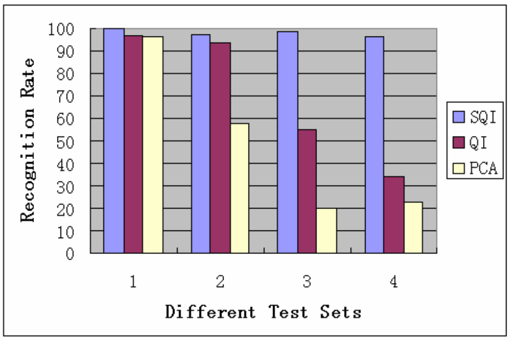


Figure 6. Recognition Results on **Yale B**

5. Conclusion

We have introduced a new algorithm, self-quotient image (**SQI**), for robust face recognition under various lighting conditions. Illumination invariant and variant properties of self-quotient algorithm are analyzed according to the Lambertian model. Though SQI has similar form as QI, it needs only one face image for implementation and no alignment in SQI procedure is needed. This algorithm has special ability of de-shadow. The experiment results show that the **SQI** method can significantly improve the recognition rate to face images under different lighting conditions.

Reference

[1] P. N. Belhumeur, David J. Kriegman, "What is the set of Images of an Object Under All possible Lighting Conditions?", IEEE conf. On Computer Vision and Pattern Recognition, 1996

[2] Athinodoros S. Georghiades and Peter N. Belhumeur, "Illumination cone models for Faces recognition under variable lighting ", CVPR, 1998

[3] Athinodoros S. Georghiades and Peter N. Belhumeur, "From Few to many: Illumination cone models for face recognition under variable lighting and pose", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 6, pp 643-660, 2001

[4] Amnon Shashua, and Tammy Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations", Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp129-139, 2001

[5] T. Riklin-Raviv and A. Shashua. "The Quotient image: Class based recognition and synthesis under varying illumination". In Proceedings of the 1999 Conference on Computer Vision and Pattern Recognition, pages 566--571, Fort Collins, CO, 1999.

[6] Ravi Ramamoorthi, Pat Hanrahan, " On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object", J. Opt. Soc. Am., Vol. 18, No. 10, 2001

[7] Ravi Ramamoorthi, "Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 10, 2002-10-21

[8] Ravi Ramamoorthi and Pat Hanrahan, "An Efficient Representation for Irradiance Environment Maps", SIGGRAPH 01, pages 497--500, 2001

[9] Ronen Basri, David Jacobs, "Lambertian Reflectance and Linear Subspaces", NEC Research Institute Technical Report 2000-172R

[10] Ronen Basri and David Jacobs, Lambertian Reflectance and Linear Subspaces, IEEE Transactions on Pattern Analysis and Machine Intelligence, forthcoming

[11]Terence Sim, Takeo Kanade, "Illuminating the Face", CMU-RI-TR-01-31, Sept. 28, 2001

[12]Yael Adnin, Yael Moses and Shimon Ullman, "Face recognition: The problem of compensating for changes in illumination direction", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, 1997， pp712-732.

[13] Daniel J. Jobson, Zia-ur Rahman, and Glenn A. Woodell, " Properties and Performance of a Center/Surround Retinex", IEEE Transactions on Image Processing, Vol. 6, No. 3, 1997, pp 451-462

[14] Daniel J. Jobson, Zia-ur Rahman, and Glenn A. Woodell, " A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes", IEEE Transactions on Image Processing, Vol. 6, No. 7, 1997, pp 965-976

[15] Ralph Gross, Vladimir Brajovie, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognitoin", 4th International Conference on Audio and Video Based Biometric Person Authentication, pp. 10-18, 2003

[16]R. Epstein, A. L. Yuille and P.N. Bellumeur, " Learning Object Reorientations form Lighting Variation", in Object Rep. in Computer Vision II (J. Ponce, A. Zisserman, and M. Hebert, eds.), pp.179--199, Springer-Verlag, 1996

[17] P. Hallanan, "A Low-Demensional Representation of Human Faces for Arbitary Lighting Conditions", IEEE Conf. On Computer Vision and Pattern Recognition, pp 995-99, 1994

[18] R. Epstein, P. Hallanan, A. L. Yuille, "5±2 Eigenimages Suffice: An Empirical Investigation of Low-Dimensional Lighting Models", IEEE Conf. Workshop on Physics-Based Vision, pp 108-116, 1995.

[19] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database**",** Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, May, 2002

[20]E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision,", in Proc. Nat. Adad. Sic., vol. 83