Self Quotient Image for Face Recognition

[1]Haitao Wang, [2]Stan Z. Li, [1]Yangsheng Wang, [3]Jianjun Zhang

[1]Institute of Automation, Chinese Academy of Sciences; [2]Miscrosoft Research Asian; [3]Media School, Bournemouth University

ABSTRACT

The reliability of facial recognition techniques is often affected by the variation of illumination, such as shadows and illumination direction changes. In this paper, we present a novel framework, called the self-quotient image, for the elimination of the lighting effect in the image. Although this method has a similar invariant form to the quotient image by Shashua etc [1], it does not need the alignment and bootstrap images. Our method combines the image processing technique of edge-preserved filtering with the Retinex applications of by Jobson, et al [2] and Gross and Brajovie [3]. We have analyzed this algorithm with a 3D imaging model and formulated the conditions where illumination-invariant and –variant properties can be realized, respectively. A fast anisotropic filter is also presented. The experiment results show that our method is effective in removing the effect of illumination for robust face recognition.

1. INTRODUCTION

The illumination problem has been considered as the one of most difficulties in face recognition and has received much attention in recent years. It is well known that image variation resulting from light change is more significant than that from different personal identities.

In recent years, many algorithms have been developed, such as the Illumination Cone [4], Spherical Harmonic subspace [5] methods.

Compared with these algorithms, the Quotient Image (**QI**) [1] is both simple and practical. It has been proven that the quotient image, which is an image ratio between a test image and a linear combination of three images illuminated by non-coplanar lights, depends only on the albedo information, and therefore is illumination free. However, the QI method makes a number of assumptions, including the facial shape, the absence of shadows, and the alignment between the images, which significantly limits its application. Our method does not make any assumptions.

Nayar and Rolle [6] also advanced one kind of image ratio. This kind of QI is the ratio of an image with its neighboring points. Under the assumption of Lambertian Model, they deduced that this QI is the ratio of reflectance coefficients, which is illumination free. Our Self Quotient Image (**SQI**) algorithm has a similar form, but we analyze the invariant properties of **SQI** for all cases, including shading region, shadow region and edge region. Furthermore we introduce a simple edge-preserving filter to produce a smoother version of the original image.

Jacobs etc [7] introduced another kind of **QI**, which is the ratio of two images. They showed that for a point illumination source and objects with Lambertian reflectance, the ratio of two images from the same object is simpler than the ratio of images from different objects. Similar to Nayar and Rolle's approach [6], Jacobs's method only considers the Lambertian model without shadow and assumes the surface of the object is smooth.

The Retinex theory proposed by Land [8] deals with illumination effects on images. Jobson, et al [2] presented a multi-scale version of Retinex method for high quality visual displays of high dynamic range image on low dynamic devices, such as printers and computer screens. This is closely related to the illumination issue. More recently Gross and Brajovie [3] presented an anisotropic version of Retinex for illumination normalization.

Unlike their reflectance-illumination imaging model, we theoretically analyze our algorithm with a 3D imaging model and formulate the illumination-invariant and –variant properties of this method in this paper.

2. SQI FRAMEWORK

The Lambertian model can be factorized into two parts, the intrinsic part and the extrinsic part:

$$\begin{aligned} I(x,y) &= \rho(x,y)n(x,y)^T \bullet s \\ &= F(x,y) \bullet s \end{aligned} \tag{1}$$

where $\boldsymbol{\rho}$ is albedo and $\boldsymbol{n}$ is the surface normals.

In the above, $\boldsymbol{F=\rho n^T}$ depends on the albedo and surface normal of an object and hence is intrinsic. It is $\boldsymbol{F}$ that represents the identity of a face. $\boldsymbol{s}$ is the illumination and is an extrinsic factor. Current appearance-based methods, including PCA, learn a representation from

0-7803-8554-3/04/$20.00 ©2004 IEEE.

image I and hence mix the intrinsic factor for the identity with the extrinsic factor. This is one of the main problems in accurate face recognition. Separating the two factors and removing the extrinsic factor are a key to achieving robust face recognition.

Our SQI method, as shown in Figure 1, has two main steps: (1) illumination estimation and (2) the illumination effect subtraction.

First, the extrinsic factor is estimated and a synthesized image is generated. The synthesized face image has the same illumination and shape as the input but a different albedo. Then the illumination is normalized by taking the difference between the logarithms of the input and the synthesized images. Because the synthesized image has the same 3D shape and illumination as the original one, the normalized image is ($\log\rho_0 - \log\rho_1$), where ρ_0 and ρ_1 are the albedo maps of the input and synthesized images, respectively; and is therefore illumination-free.

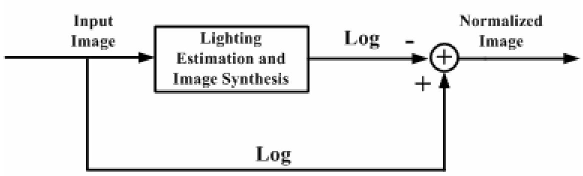


Figure 1 Illumination Normalization

The logarithm is necessary because if the subtraction were between the two images directly, the result Q would be $Q = I - \hat{I} = (\rho_0 - \rho_1)n^T s$, which would still be illumination dependent.

3. SELF QUOTIENT IMAGE

3.1. Definition

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

Definition 1: Self-Quotient Image. The Self-Quotient image Q of image I is defined by

$$Q = \frac{I}{\hat{I}} = \frac{I}{F * I} \tag{2}$$

where $\hat{I}$ is the smoothed version of I, 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 a kind of quotient image derived from the image I itself rather than images of a different person as in QI.

3.2. Analysis

In the following analysis of the **SQI**, we consider three cases with different shapes and shadow conditions.

Case 1: In regions without shadow and with small surface normal variation.

In this case, $n^T(u,v)s \approx C_1$, where C_1 is a constant. Then we have

$$Q = \frac{I}{\hat{I}} \approx \frac{\rho C_1}{(\rho * F)C_1} = \frac{\rho}{\rho * F} \tag{3}$$

In this case, $\boldsymbol{Q}$ is approximately illumination free and depends only on the albedo of the face. Equation (2) 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 surface normal variation.

In this case, $n^T(u,v)s$ is not a constant. The SQI is

$$Q = \frac{I}{\hat{I}} = \frac{\rho n^T s}{F * (\rho n^T s)} \tag{4}$$

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

Case 3: In shadow regions.

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

$$\begin{aligned} I(u,v) &= n(u,v)^T \sum_{i=1}^{\infty} s(u,v)_i \\ &= \sum_{i=1}^{\infty} n(u,v)^T s(u,v)_i = C_2 \end{aligned} \tag{5}$$

where C_2 is a constant. Therefore, $I(u,v)$ in shadow regions can be written as $I \approx C_2\rho$.

Then we have

$$Q = \frac{I}{\hat{I}} \approx \frac{C_2\rho}{(C_2\rho) * F} = \frac{\rho}{\rho * F} \tag{6}$$

As in case 1, **SQI** in this kind of regions is also illumination-free; in other words, the **SQI** removes the shadow effect, shown in Fig. 2.

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

$$I = \rho n^T S = \rho n^T \sum_{i=1}^{L} s_i \tag{7}$$

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

Figure 2 Light normalization using SQI

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

For face recognition, if we can ensure that the filter's kernel size is small enough compared with the variation of the face surface normal, the self-quotient image will be illumination free as previously analyzed. However, when the filter's kernel size is too small, $\boldsymbol{Q}$ will approach one and the albedo information is lost.

The advantages of the self-quotient method as opposed to the original quotient image is summarized as follows: (1) The alignment between image I and its smoothed version $\widehat{I}$ is automatically perfect, and hence it does not need an alignment procedure. (2) No training images are needed for the estimation of the illumination direction because the illumination fields of I and $\widehat{I}$ are similar. (3) the self-quotient image is good at removing shadows, whereas in the previous approaches [1,4,5,7], either ignored the shadow problem or solved it by a complex 3D reconstruction. (4) Lighting sources can be any type.

In the implementation, we use the multi-scale technique to improve algorithm robustness. In particular, we choose different kernel sizes to suit different regions.

3.3. Algorithm

Though Jobson's [2] filter is simple, it is isotropic and creates the undesirable halo effects around the edge region. Gross's [3] anisotropic filter, which can reduce the halo effect, uses an iterative procedure. For real time applications, this method is too computationally expensive. Let us consider a weighed Gaussian filter, given by

$$F = WG \tag{8}$$

where $\boldsymbol{W}$ is the weight and $\boldsymbol{G}$ is the Gaussian kernel.

Let $\boldsymbol{\Omega}$ be the convolution region. We divide $\boldsymbol{\Omega}$ 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 $\tau = Mean(I_\Omega)$. For the two sub-regions, $\boldsymbol{W}$ has the following corresponding values.

$$W(u,v) = \begin{cases} 0 & \mathrm{I(u,v)} \in \mathrm{M}_2 \\ 1 & \mathrm{I(u,v)} \in \mathrm{M}_1 \end{cases} \tag{9}$$

If the convolution image region is smooth, i.e. little gray value variation (non-edge regions), there is also little difference between the smoothing the whole region and smoothing part of the region. If there is a large gray value variation in the convolution region, i.e. in an 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.

In the application, we adopt this filter with a multi-scale version, i.e. the linear combination of the output of different Gaussian $\boldsymbol{\sigma}$.

The main difference between the Quotient Image [1] and the **SQI** is in the illumination estimation. In **QI**, they estimate the single point lighting source by coefficients of the three non-collinear illuminated images. In **SQI**, we do not estimate the illumination directly and the same illumination condition is implied in the smoothed version.

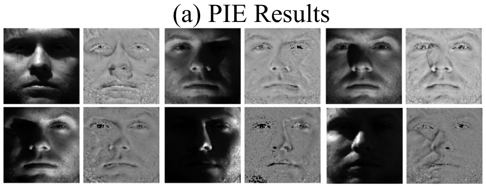
4. EXPERIMENTS AND DISCUSSION

We test our algorithm on two face databases, Yale B face database [3] and CMU PIE face database [9]. Only the frontal images with lighting variation are selected and manually cropped. We also divide the Yale B in 4 Sets with increasing lighting angles. The classifier used in experiments is the nearest neighbor with the correlation coefficients as its distance measurement.

We use only one frontally illuminated image as a template in Yale B for the face recognition experiment. In PIE, we adopt each image as a template and the rest as test images. All the images have been converted into self-quotient images before undertaking the recognition process. 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 3 shows some results of the **SQI** based illumination normalization. We can see that the convolution based anisotropic filtering is very effective in smoothing the noisy images without blurring the step edges. Shadows are removed.

(a) PIE Results



(b) Yale B

Figure 3 Example results of **SQI for** illumination elimination

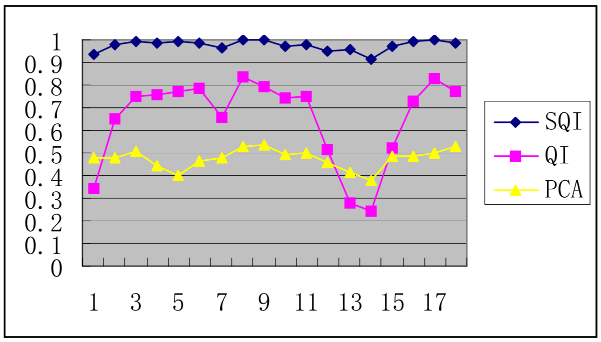


Figure 4 Recognition results on PIE.

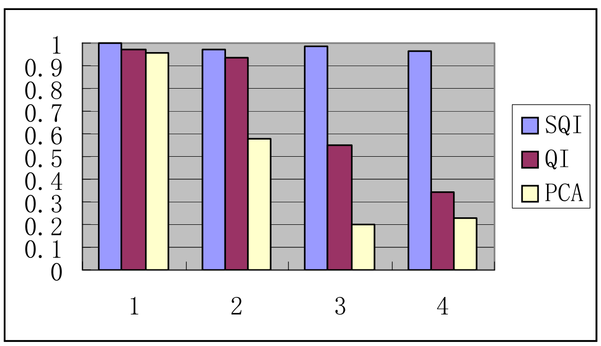


Figure 5 Recognition results on Yale B.

For the PIE data set, the leave-one-out scheme is used, i.e. each image is used as a template in turn and the others as test images. The results are compared in Figure 4 for the 20 different leave-one-out partitions. For the Yale B data set, the images are divided into 4 subsets according to their increasing illumination angles, and only the frontally illuminated images are used as the templates. The results are shown in Figure 5 for the 4 different data sets.

Compared with the **PCA** and the original **QI**, our **SQI** can significantly improve the recognition rates both for CMU PIE and Yale B face databases.

5. COCLUSION

A generalized **SQI** framework is presented. This unified framework explains the essence of previous **QI** [1], Retinex-based [3] and image ratio-based [6,7] algorithms without any assumption on illumination conditions and shadows. Under this framework, we first define the self-quotient image as a new illumination invariant representation of face images. Then we analyze its illumination invariant and variant properties using the Lambertian model. We also developed a novel anisotropic filter in the implementation of the self-quotient algorithm. The experiment results show that our method can significantly improve the recognition rates of face images under different lighting conditions.

REFERENCES

[1] 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

[2]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

[3] 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

[4]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

[5]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, pp 2448-2459 ,2001

[6]S. K. Nayar and R. M Rolle, "Reflectance Based Object Recognition", International Journal of Computer Vision, 17(3):219--240, 1996.

[7]D. Jacobs, P. Belhumeur and R. Basri, "Comparing Images Under Variable Illumination", IEEE Conference on Computer Vision and Pattern Recognition, pp. 610-617, 1998.

[8]E. Land, "The Retinex Theory of Color Vision", Sci. Amer. 237:10-128, 1977

[9]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, 2002