8

Illumination Normalization using Quotient Image-based Techniques

Masashi Nishiyama, Tatsuo Kozakaya and Osamu Yamaguchi
Corporate Research & Development Center, Toshiba Corporation
Japan

1. Introduction

This chapter focuses on correctly recognizing faces in the presence of large illumination variation. Our aim is to do this by synthesizing an illumination normalized image using Quotient Image-based techniques (Shashua et al., 2001, Wang et al., 2004, Chen et al., 2005, Nishiyama et al., 2006, Zhang et al., 2007, and An et al., 2008). These techniques extract an illumination invariant representation of a face from a raw facial image.
To discuss the variation of facial appearances caused by illumination, the appearances are classified into four main components: diffuse reflection, specular reflection, attached shadow and cast shadow (See Figure 1) as described in (Shashua, 1999).

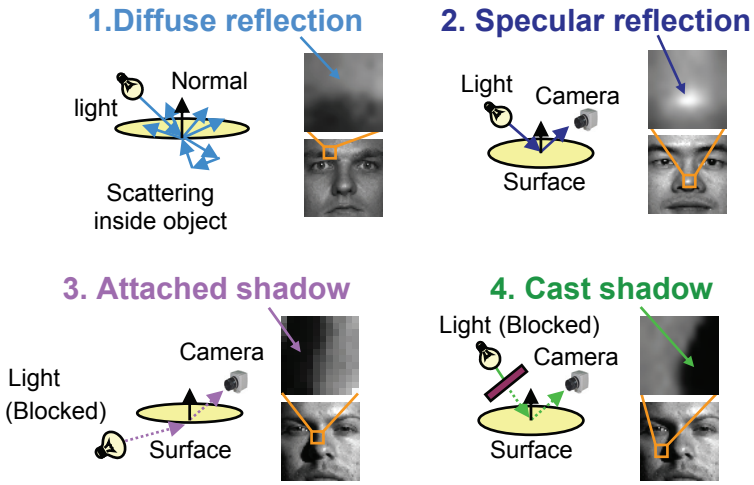

Fig. 1. Facial appearance classified into four main components: diffuse reflection, specular reflection, attached shadow and cast shadow. Diffuse reflection occurs when the incident light is scattered by the object. The pixel value of diffuse reflection is determined by albedo that is invariance (See Equation (2)). Specular reflection occurs when the incident light is cleanly reflected by the object. Attached shadow occurs when the object itself blocks the incident light. Cast shadow occurs when a different object blocks the incident light.

The Self-Quotient Image (SQI) (Wang et al., 2004) of a pixel is defined as the ratio of the albedo of that pixel to a smoothed albedo value of local pixels. The SQI however is neither synthesized at the boundary between a diffuse reflection region and a shadow region, nor at the boundary between a diffuse reflection region and a specular reflection region. This problem is not addressed in the Total Variation Quotient Image (Chen et al., 2005, An et al., 2008) and the Morphological Quotient Image (Zhang et al., 2007). Determining the reflectance type of an appearance from a single image is an ill-posed problem.
Our technique exploits a learned statistical model to classify the appearances accurately. Our statistical model is defined by a number of basis images representing diffuse reflection on a generic face. These basis images are trained on faces that are completely different to those later registered for identification. Synthesized images using our model are called the Classified Appearance-based Quotient Images (CAQI). The effectiveness of CAQI is demonstrated on the Yale Face Database B and a real-world database (see section 4). We compare CAQI with SQI, Retinex (Jobson et al., 1997) images produced by an equation similar to SQI, and conventional histogram equalization for illumination normalization. From these results, we see a significant improvement by classifying appearances using our Quotient Image-based technique.

1.1 Background

A method for synthesizing illumination normalized images without an explicit appearance model has been proposed in (Belhumeur et al., 1997, Nishiyama et al., 2005), where the variation caused by illumination is regarded as an intra-class variation. An illumination normalized image is synthesized by projection onto a subspace that is insensitive to the intra-class variation due to illumination, but sensitive to the interclass variation representing changes in appearance between individuals. However, in the case of different lighting conditions amongst training images for the subspace, an illumination normalized image cannot be synthesized since the subspace cannot estimate a novel appearance that is not included in the training images.
In order to apply an explicit appearance model for illumination, the basis images method has been proposed in (Shashua et al., 2001, Okabe et al, 2003), which is suitable for modelling the variation in facial appearance caused by diffuse reflection. This model is advantageous in that a registration system stores only a few images and fitting a face image to the model is simple. Shashua et al. have proposed the Quotient Image technique (QI) (Shashua et al., 2001), which is the ratio of albedo between a face image and linear combination of basis images, for each pixel. This ratio of albedo is illumination invariant. However, the QI assumes that a facial appearance includes only diffuse reflection. Okabe et al. synthesized a normalized image consisting of diffuse reflection and attached shadow, removing specular reflection and cast shadow from a face image (Okabe et al, 2003). This method estimates diffuse reflection and attached shadow using basis images and random sample consensus. Basis images are generated from three face images acquired under fixed pose and a moving point light source (Shashua, 1999), or from four face images acquired under a moving pose and a fixed point light source (Nakashima et al., 2002). Therefore, these methods (Shashua et al., 2001, Okabe et al, 2003) require multiple face images for each individual in the registration stage of training. This requirement is an important limitation for practical applications. In the case that there is only one training image available per person, these methods cannot be applied.

2. Obtaining invariance with respect to diffuse reflection

2.1 Self-Quotient Image (SQI)

The SQI technique (Wang et al., 2004) has been proposed for synthesizing an illumination normalized image from a single face image. The SQI is defined by a face image $I(x, y)$ and a smoothed image $S(x, y)$ as

$$Q(x,y) = \frac{I(x,y)}{S(x,y)} = \frac{I(x,y)}{F(x,y) * I(x,y)}, \tag{1}$$

where $F(x, y)$ is a smoothing filter; $*$ is the convolution operation. In the case that a smoothing filter is an isotropic Gaussian filter $G(x, y)$, equation (1) is equivalent to the center/surround retinex transform described in (Jobson et al., 1997). However, in the case of the SQI technique, an anisotropic, weighted Gaussian filter $W(x, y)G(x, y)$ is used.

As described in (Wang et al., 2004), $Q(x, y)$ is illumination invariant if certain assumptions are met (see assumptions (a) and (b)). The diffuse reflection is defined by the Lambertian model as

$$i = \max(al\mathbf{n}^T\mathbf{s}, 0), \tag{2}$$

where i is pixel value of $I(x, y)$; a and $\mathbf{n}$ are the albedo and the normal of the object surface; l and $\mathbf{s}$ are the strength and direction of the light source. The lengths of $\mathbf{n}$ and $\mathbf{s}$ are normalized. Attached shadow appears in the case where the dot-product between $\mathbf{n}$ and $\mathbf{s}$ is negative. We make two assumptions in a local region: that (a) l, $\mathbf{n}$, and $\mathbf{s}$ are uniform and (b) all observed appearance is diffuse reflection. Then the ratio of albedo is extracted as

$$Q(x,y) = \frac{a(x,y)l\mathbf{n}^T\mathbf{s}}{F(x,y) * a(x,y)l\mathbf{n}^T\mathbf{s}} = \frac{a(x,y)}{F(x,y) * a(x,y)}. \tag{3}$$

Under multiple light sources, the ratio of albedo is also obtained using additivity $\mathbf{s} = \sum_{k=1}^{n} \mathbf{s_k}$.

2.2 Problems caused by effects other than diffuse reflection

Equation (3) works well for a local region such as Figure 2 (i) that includes the low albedo region of the eyebrow and the middle albedo region of the forehead. The assumptions of section 2.1 are valid in this local region where only diffuse reflection is observed. However, the assumptions are violated in Figure 2 (ii) and (iii) where cast shadow, or specular reflection, is partially observed. In the specular reflection region, diffuse reflection pixel values cannot be extracted because the pixel values are saturated. Then, we need to estimate diffuse reflection pixel values from the pixels that are in the specular reflection region. In a cast shadow region, light parameters, l and s, are different from those of diffuse reflection because there is an obstruction between the light source and the facial region of interest. The same is true of the attached shadow regions. In these regions, diffuse reflection can be sometimes observed because the region is illuminated by another light source e.g. ambient light. Then, we need to determine a local region that is all subject to the same lighting conditions for equation (3) to work well.

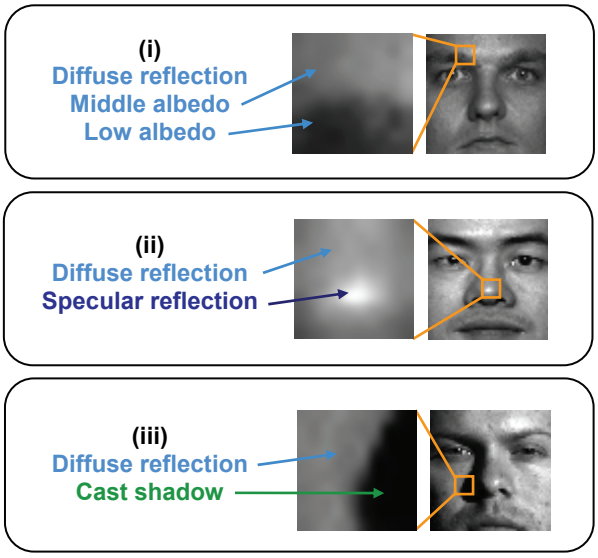

Fig. 2. Examples of different appearances in local regions. Reflections other than diffuse reflection are also observed in local regions (ii) and (iii). This means that, using the QI technique, an illumination invariant feature cannot be extracted from these local regions.

2.3 Weight function for synthesizing the SQI

The weight function $W(x, y)$ of the SQI is designed for each pixel to prevent halo effects either side of the edges in the image. The weight function divides a local region into two subregions M_1 and M_2. If $I(x, y) \in M_1$ then $W(x, y) = 1$ else $W(x, y) = 0$. The subregion is determined by a threshold τ that is the average of pixel values in the local region. The subregion with the larger number of pixels is set to M_1, the other is set to M_2. However, the weight function is unable to discriminate between shadow and low albedo regions, e.g. eyes and eyebrows. Illumination invariant features can only be extracted from a local region containing an edge between a middle albedo region and a low albedo region such as Figure 2(i). This inability of the SQI to discriminate between shadow and low albedo region results important information representing identity being neglected. Another problem is that a region is not simply divided where pixel values change smoothly, e.g. soft shadow is observed on the edge of the cast shadow.

3. Classified Appearance-based Quotient Image (CAQI)

In this section we introduce a new method for extracting the ratio of albedo from a face image that includes specular reflection and shadow. In this method the weight function for the weighted Gaussian filter is calculated by classifying the appearance of each pixel.

3.1 Classification of appearance caused by illumination using photometric linearization

To classify surface appearance into diffuse reflection, specular reflection, attached shadow and cast shadow, we utilize photometric linearization (Mukaigawa et al., 2001). Photometric

linearization transforms a face image into a linearized image consisting of only diffuse reflection. A linearized image $\widetilde{I}(x,y)$ is defined by a model (Shashua, 1999) in which arbitrary images with only diffuse reflection are represented by a linear combination of three basis images $I_i(x,y)(i=1,2,3)$ taken under point light sources in linearly independent directions as

$$\widetilde{I}(x,y)=\sum_{i=1}^{3}c_i I_i(x,y). \tag{4}$$

To estimate coefficients c_i from a face image including specular reflection and shadow, photometric linearization uses random sampling of pixels. By random sampling we calculate candidates of c_i. The final c_i is determined from the distribution of candidates by iterating between outlier elimination and recomputation of the mean. The appearance is classified using a difference image $I'(x,y)$, defined as

$$I'(x,y)=I(x,y)-\widetilde{I}(x,y). \tag{5}$$

As in (Mukaigawa et al., 2001) a pixel having a negative value in $I'(x,y)$ is classified as cast shadow. A pixel having a value greater than the threshold in $I'(x,y)$ is classified as specular reflection. A pixel having a negative value in $\widetilde{I}(x,y)$ is classified as attached shadow. To estimate diffuse reflection for a pixel that has specular reflection, we replace the pixel value $I(x,y)$ with $\widetilde{I}(x,y)$.

3.2 Generation of basis images

Since we do not acquire basis images $I_i(x,y)$ for each individual in the training stage, we generate basis images using different individuals from those in the training stage. For this purpose, we acquire images under fixed pose and a moving point light source. These images do not exhibit specular reflection or shadow. The point light source is moved so as to make each basis image linearly independent. We apply Singular Value Decomposition (SVD) using all the acquired images. The vectors, selected in descending order of singular value, are the basis images. We assume that the basis images represent the diffuse reflection on a generic face. However, retaining the first three basis images, the estimated c_i in equation (4) has an error induced by the difference between a generic face and an individual face. For fitting basis images to various individuals, we select more than four vectors in descending order of singular value. These vectors represent the principal component of the variation in appearance in a diffuse reflection of the individuals.

3.3 Calculation of the weighted function using classified appearance

To calculate the weight function for the weighted Gaussian filter, we utilize a difference image $I'(x,y)$ representing the difference of the appearance. We aim to extract the ratio of albedo from a local region in which similar appearance is observed. For example, if the center of a local region is classified as diffuse reflection, we give a large weight to diffuse

reflection of the surrounding area of the center and small weight to areas in shadow. The weight function $W(x, y)$ is defined by comparing the center of a Gaussian filter with its surrounding area as follows:

$$W(x,y) = \frac{1}{1+\alpha\left|I'(x,y) - I'(x_0,y_0)\right|}, \tag{6}$$

where (x_0, y_0) is the center pixel of the Gaussian filter; α is constant $(\alpha > 0)$. If $I'(x, y)$ is greater than $I'(x_0, y_0)$, then $I'(x, y)$ has a different appearance to $I'(x_0, y_0)$, and a small weight is given to $I'(x, y)$.

3.4 Algorithm for synthesizing the CAQI

We now explain the algorithm for synthesizing the Classified Appearance-based Quotient Image (CAQI). The flow diagram of the algorithm is shown in Figure 3. First, a face image is aligned from the positions of the pupils and the nostrils. Next, appearance is classified using photometric linearization. Then, the specular reflection is replaced with the estimated diffuse reflection.

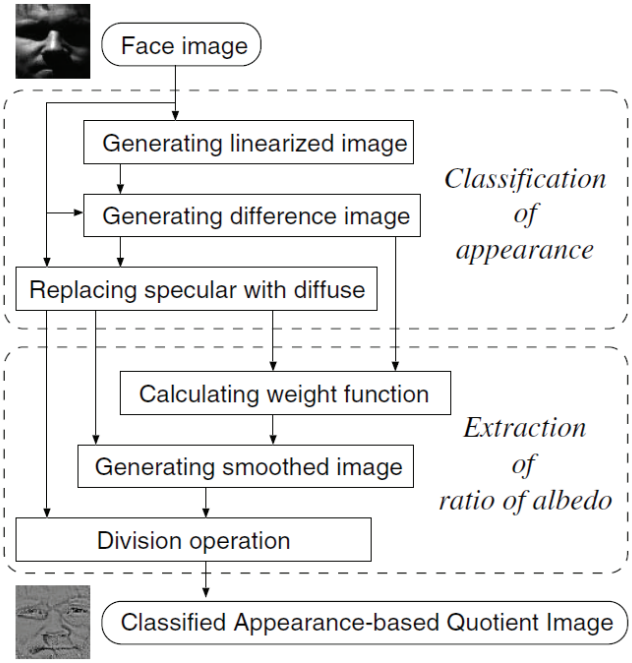

Fig. 3. Flow of synthesizing the CAQI.

For calculating the ratio of the albedo, we need to determine the size of a local region in order to obtain a uniform surface normal $\mathbf{n}$. However, searching for a suitable size is difficult since there is an ambiguity in the computation of $\mathbf{n}$ for basis images that are generated by unknown l and $\mathbf{s}$ (Hayakawa, 1994). Therefore we use multiple sizes as in

(Jobson et al., 1997, Wang et al., 2004). The size is defined by the standard deviation σ of the Gaussian filter used in the smoothing operation. We calculate the weight function $W_j(x,y)$ for each $\sigma_j (j=1,...,N)$. Finally, the illumination normalized image $Q(x,y)$ is synthesized as

$$Q(x,y)=\sum_{j=1}^{N} f\left(\frac{I(x,y)}{W_j(x,y)G_j(x,y)*I(x,y)}\right). \tag{7}$$

Note that we assume $\int\int W_j(x,y)G_j(x,y)dxdy=1$. As in (Jobson et al., 1997, Wang et al., 2004) we use f in equation (7). The function f prevents asymptotic approach to infinity when the term $W_j(x,y)G_j(x,y)*I(x,y)$ tends to zero in shadowed regions.

4. Empirical evaluation

4.1 Performance for varying illumination

4.1.1 Experimented conditions

To illustrate the performance of our method, we have conducted face identification experiments using the Yale Face Database B (Georghiades et al., 2001). The database consists of images taken under 64 different lighting conditions, which are divided into 5 subsets. In Figure 4, we show examples of face images in each subset. We used 640 images in total taken of 10 individuals in a frontal pose. We located a 64×64 pixels face image from the positions of the pupils and the nostrils, obtained manually. In our experiments we use a single face image of each individual for training, where the lighting conditions are the same for all subjects. This is different from (Wang et al., 2004) which used multiple images for training. The images taken under the remaining 63 lighting conditions are used for testing.

For the generation of basis images, we used the CMU PIE (Sim et al., 2003) consisting of different individuals from the Yale Face Database B. We selected images taken from a frontal pose (c27) under light sources (f06, f07, f08, f09, f11, f12, f20, f21). We applied SVD to all the images of 68 individuals and selected the 7 basis images shown in Figure 5.

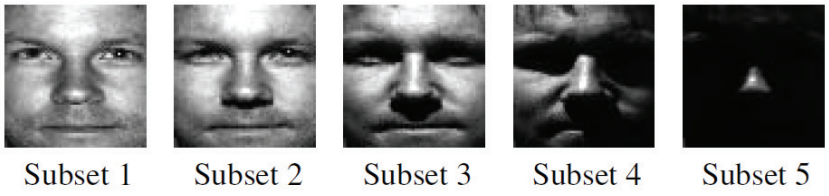

Fig. 4. Examples from the Yale face database B.

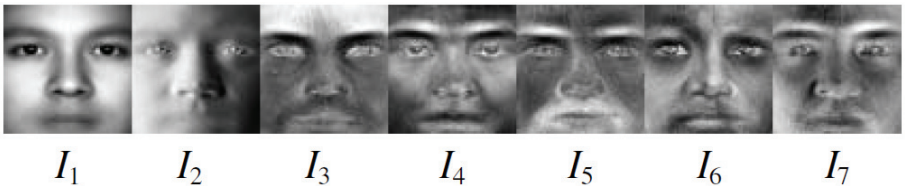

Fig. 5. Basis images generated from the 68 CMU-PIE individuals under 8 lighting conditions.

4.1.2 Identification performance

We compared the performance of the CAQI with the gray-scale image (GS), the histogram equalized image (HE), the Multi-scale Center/Surround Retinex image (MSR) (Jobson et al., 1997) and the SQI (Wang et al., 2004). For GS, we used the face image directly. For HE, we used the histogram equalization of face image. For MSR, SQI and CAQI, we used multiple Gaussian filters, σ_i were set to 1.0, 2.0, 3.0, and f in equation (7) was chosen as the logarithmic function. For CAQI, we sampled 20,000 candidates to estimate c_i and α was set to 0.1. The synthesized image was transformed to a vector by raster-scanning of the image and the length of the vector was normalized. Then, we calculated the similarity between the training vector and the testing vector using normalized correlation.

Table 1 shows the evaluation results for each method in terms of the correct recognition rate (RR). RR is the probability that a face image of the right person is accepted correctly when using nearest neighbor classification. We show the average of the RR for each subset in the table. The performance of the MSR is superior to that of the SQI. Heusch et al. reported a similar result for a different database (Heusch et al., 2005). For this reason, we infer that the weight function for the SQI causes the problem that the center pixel of Gaussian filter is not included in the subregion M_1. Since the appearance differs between the center pixel and the subregion, the assumptions for equation (3) are invalid. From the table we can see that the performance of CAQI is superior to that of the others. In particular, the performance in subset 5 where faces are mostly shadowed is improved significantly.

Subset (Training)	Method				
	GS	HE	MSR	SQI	CAQI
1	68	71	93	87	96
2	64	68	91	82	95
3	54	58	82	71	88
4	39	43	79	68	84
5	25	42	79	75	91

Table 1. RR (%) on the Yale Face Database B. Images taken under one lighting condition were used for training set. Images taken under 63 different lighting conditions were used for test set.

4.1.3 Comparison of synthesized images

Figure 6 shows the synthesized center/surround retinex image, SQI and CAQI. To synthesize these images we used a single σ = 1.0. In the case of the synthesized images in subset 4, features hidden by shadow around the eyes in the gray-scale image appear in the center/surround retinex image, SQI, and CAQI. It can clearly be observed that the influence of cast shadows is reduced in the CAQI. This is particularly evident at the edge of the shadowed region and in the region of specular reflection on the nose.

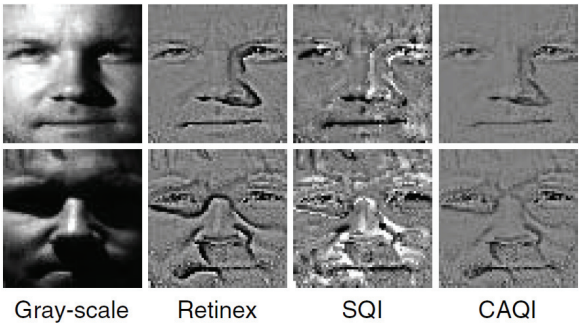

Fig. 6. Examples of an image synthesized by each method. (The upper row is subset 3 and the lower row is subset 4)

4.1.4 Identification performance in the case of basis images for each individual

We compared the identification performance using basis images of each individual (CAQI-same) and using basis images of others (CAQI-other). In CAQI-other, we used basis images of CMU-PIE shown in Figure 5. In CAQI-same, basis images for each individual in a training set were generated using 7 face images in subset 1. We selected three basis images in descending order of singular value. Images taken under one lighting condition in subset 1 were used for training set. Images taken under 63 different lighting conditions were used for the test set. Table 2 shows the evaluation result in terms of RR. CAQI-other and CAQI-same are superior to the SQI. However, CAQI-same is superior to CAQI-other. For this reason, we presume that the albedo and the surface normal of faces in the test set are different from those for individuals included in CMU-PIE. To increase performance further, we are considering developing a method of deforming basis images from generic basis images without requiring more than one training image per person.

Method	Subset (Input)				
	1	2	3	4	5
SQI	100	97	91	76	70
CAQI-other	100	97	97	93	93
CAQI-same	100	99	99	95	94

Table 2. Comparison of identification performance using basis images of each individual versus basis images of others. Images taken under one lighting condition in subset 1 were used for the training set.

4.2 Performance assessment on a real-world database

We also evaluated the methods using a database collected under two lighting conditions for 100 individuals. We assumed a practical application of face recognition, namely to passport-based identification. The lighting conditions are (I) no shadow on the facial appearance by using a flash attached to the camera, and (II) face shadowed by a single spotlight mounted

on the ceiling as shown in Figure 7. We collected a single face image for each individual and each lighting condition when the ceiling fluorescent lamps were on. The parameters of section 4.1 remained unchanged. We show the RR (%) in Table 3 for the condition that (I) is the training set and (II) is the testing set, and vice versa.

We can see that the method using the CAQI is superior to the other methods. In particular, the CAQI is effective when images containing cast shadow regions, as in (II), are used as training examples. However, an error in the photometric linearization arose since a single set of c_i does not fully represent the diversity of illumination under multiple light sources. To improve the performance, we are considering developing a method of estimating multiple sets of c_i under multiple light sources.

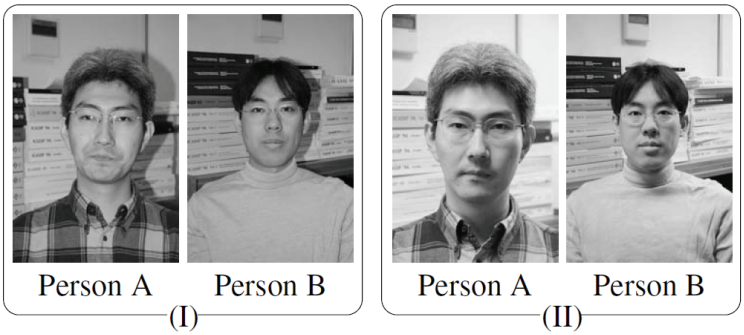

Fig. 7. Examples of a real-world database.

Training image	GS	HE	MAR	SQI	CAQI
(I)	30	16	78	65	82
(II)	33	25	73	29	80

Table 3. RR (%) on a real-world database.

5. Conclusion

This chapter discussed a method of synthesizing an illumination normalized image using Quotient Image-based techniques. Facial appearance contains diffuse reflection, specular reflection, attached shadow and cast shadow. Our Quotient Image-based technique classifies facial appearance into four components using basis images representing diffuse reflection on a generic face. Our method is able to obtain high identification performance on the Yale Face Database B and on a real-world database, using only a single image for each individual in training.

6. Acknowledgements

We would like to thank Mr. Masahide Nishiura, Dr. Paul Wyatt, Dr. Bjorn Stenger, and Mr. Jonathan Smith for their significant suggestions and comments.

7. References

An, G.; Wu, J. & Ruan, Q. (2008). Kernel TV-Based Quotient Image Employing Gabor Analysis and Its Application to Face Recognition, *IEICE Transactions on Information and Systems*, Vol. E91-D, No. 5, pp. 1573-1576

Belhumeur, P. N.; Hespanha, J. P. & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711 – 720

Chen, T.; Yin, W.; Zhou, X. S.; Comaniciu, D.; Huang, T. S. (2005). Illumination Normalization for Face Recognition and Uneven Background Correction Using Total Variation Based Image Models, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 532 – 539

Georghiades, A. S.; Belhumeur, P. N. & Kriegman D. J. (2001). 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

Hayakawa, H. (1994). Photometric stereo under a light source with arbitrary motion, *The Journal of the Optical Society of America A*, Vol. 11, No. 11, pp. 3079 – 3089

Heusch, G.; Cardinaux, F. & Marcel, S. (2005). Lighting normalization algorithms for face verification. *IDIAP-Com 05-03*

Jobson, D. J. ; Rahman, Z. & Woodell, G. A. (1997). A multi-scale retinex for bridging the gap between color images and the human observation of scenes, *IEEE Transactions on Image Processing*, Vol. 6, No. 7, pp. 965 – 976

Mukaigawa, Y.; Miyaki, H. ; Mihashi, S. & Shakunaga, T. (2001). Photometric image-based rendering for image generation in arbitrary illumination, *IEEE International Conference on Computer Vision*, Vol. 2, pp. 652 – 659

Nakashima, A.; Maki, A. & Fukui, K. (2002). Constructing illumination image basis from object motion, *Proceedings of 7th European Conference on Computer Vision*, Vol. 3, pp. 195 – 209

Nishiyama, M. ; Yamaguchi O. & Fukui, K. (2005). Face recognition with the multiple constrained mutual subspace method, *Proceedings of 5th International Conference on Audio- and Video-based Biometric Person Authentication*, pp. 71 – 80

Nishiyama, M. & Yamaguchi, O. (2006). Face Recognition Using the Classified Appearance-based Quotient Image, *Proceedings of 7th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 49 – 54

Okabe, T. & Sato, Y. (2003). Object recognition based on photometric alignment using ransac, *IEEE Proceeding Conference on Computer Vision and Pattern Recognition*, Vol.1, pp. 221 –228

Shashua, A. (1999). Geometry and photometry in 3d visual recognition, *Ph. D. Thesis*

Shashua, A. & Riklin-Raviv, T. (2001). The quotient image: Classbased re-rendering and recognition with varying illuminations, *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 129 – 139

Sim, T.; Baker, S. & Bsat, M. (2003). The cmu pose, illumination, and expression database, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1615 – 1618

Wang, H.; Li, S. Z. & Wang, Y. (2004). Generalized quotient image, *IEEE Proceeding Conference on Computer Vision and Pattern Recognition*, Vol.2, pp. 498 – 505

Zhang, Y.; Tian, J.; He, X. & Yang, X. (2007). MQI Based Face Recognition Under Uneven Illumination, *Advances in Biometrics*, Vol. 4642, pp. 290-298