Low Cost Illumination Invariant Face Recognition by Down-Up Sampling Self Quotient Image

Li Lin* and Ching-Te Chiu†

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan, R.O.C

* email: thenthey@gmail.com

† email: ctchiu@cs.nthu.edu.tw

***Abstract*—Illumination variation generally causes performance degradation of face recognition systems under real-life environments. The Self Quotient Image (SQI) method [1] is proposed to remove extrinsic lighting effects but requires high computation complexity. Therefore, we propose a low cost face recognition scheme that uses multi-scale down-up sampling to generate self quotient image (DUSSQI) to remove the lighting effects. The DUSSQI has the following advantages: (1) Remove the lighting artifacts effectively. (2) Extract different face details including texture and edges. (3) Only global operation on pixels is required to reduce computational cost. Experimental results demonstrate that our proposed approach achieves 98.58% recognition rate for extended YaleB database and 93.8% for FERET database under various lighting conditions and reduces 97.1% computational time compared to that of SQI.**

***Index Terms*—face recognition, illumination invariance, down-up sampling, quotient image**

I. Introduction

The problem of face recognition under variant lighting conditions has received a great deal of attention for its applications on smart human-machine interfaces. Illumination variation is still a challenging problem in face recognition research area. The Retinex model theory [2] represents illumination variation as the low-frequency components in a face image. According to this assumption, many methods are adopted to remove low-frequency components to retrieve illumination invariant face information. The DCT [3] transform is an example that handles the lighting condition in the frequency domain directly. The same concept is also used in the spatial domain, self quotient imaging (SQI) [1] treats the smoothed version of original face image as illumination influence on the face. When the original face image is divided by its smoothed version, the illumination invariant facial information is obtained. Other researches attempt to eliminate low-frequency components both in the preprocessing and feature extraction stages. McLauhlin *et al.* [4] combine the SQI method with illumination invariant feature, the Fourier transform magnitude for recognition.

In this paper, we focus mainly on robustness of the proposed scheme under various lighting conditions and the required computational complexity. We propose the multi-scale sampling self quotient image scheme to eliminate the illumination influence on the face image, and our approach only needs one image per-person in the gallery database. The SQI method retrieves different face detail with parameter-controlled smoothed images. When the block size and variance of the Gaussian filter are large, the obtained SQI images have details on contours such as eyes, mouth, and nose. When the block size and variance of the Gaussian filter are small, the SQI images contains delicated details such as eyebrow and pupil of the eyes. These details are embedded under different frequency components of an image. Freedman *et al.* [5] propose a method for super resolution images that adopt the down-up sampling to retrieve the low-frequency information in an image. With this assumption, we generate various scales of down-up sampling face images to retrieve different frequency information, which is more precise than the Gaussian filter used in the SQI scheme. The images processed by our proposed DUSSQI are illumination invariant. Since the DUSSQI requires only global operation on pixels so it can reduce the computational complexity significantly.

The rest of the paper is organized as follows. Section 2 presents how to generate DUSSQI images with different scales, transformation and our illumination invariant feature. Section 3 describes the experimental results. A brief conclusion is given in Section 4.

II. Down-Up Sampling Self Quotient Image (DUSSQI)

In this section, we introduce our illumination invariant face recognition approach. Fig.1 shows the flow chart of our method. First, the probe image is sampled by different down-up scale k to generate blur images that contains low frequency components. Then the self quotient image is obtained by dividing the original probe image to the blur images. The self quotient image contains different face detail depending on the scale factor k. The contrast transform is applied to enhance the contrast and adjusts the mean luminance. Therefore, using different weighting on the DUSSQI to combine these information is able to obtain an illumination invariant image. Then, the 2D Fourier transform magnitudes are used as the recognition feature. Finally, the cosine distance is applied to determine the similarity between gallery and probe face images.

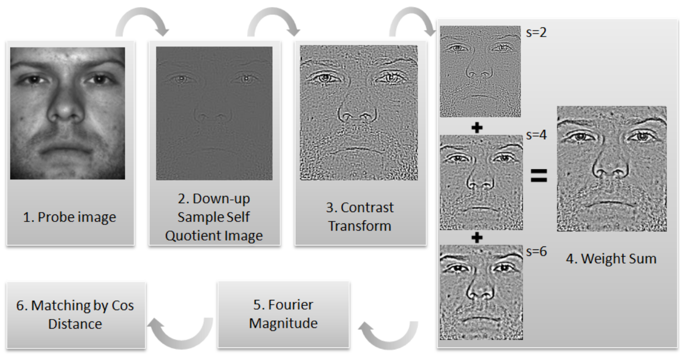


Fig. 1. The flowchart of our method with six steps. (S1) Input the probe image. (S2) Generate DUSSQI with different scales. (S3) Contrast enhancement. (S4) Combine DUSSQI with different scales. (S5) Extract feature. (S6) Similarity comparison.

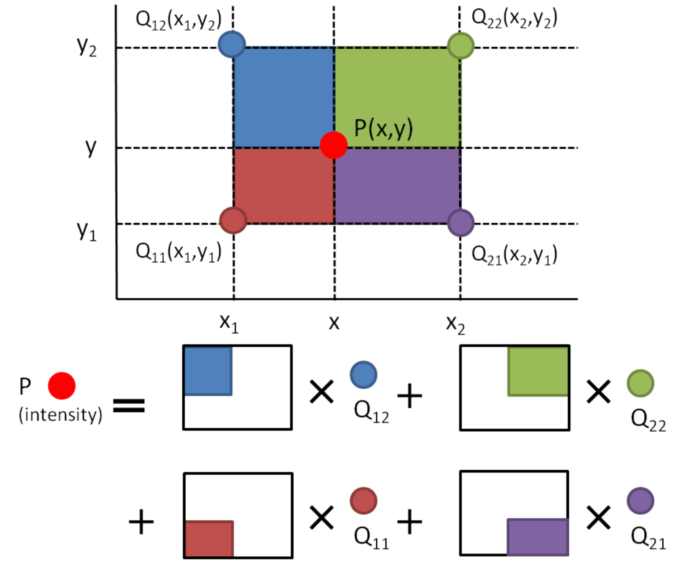
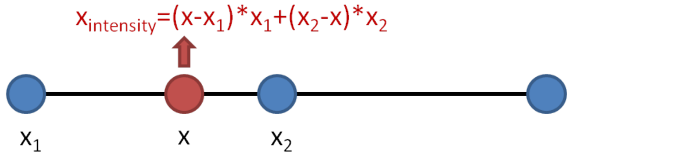
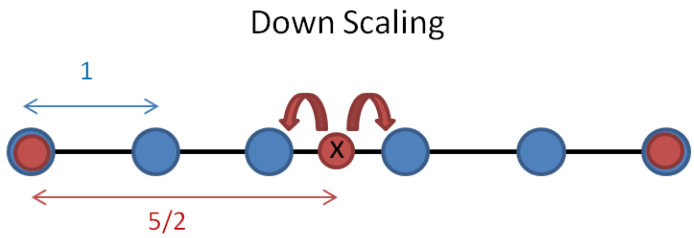


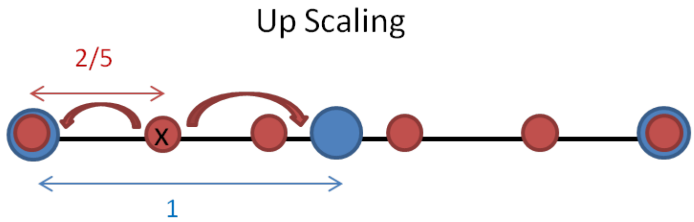
Fig. 2. Bilinear interpolation method. The intensity of interpolation point P is decided by surrounded 4 points and distance between them.



(a) The position of interpolation and original point.



(b) Example of one dimension down sampling. The $n = 6$, $k = \frac{1}{2}$, the distance between interpolation points is $\frac{5}{2}$. Red points are interpolation points, and blue points are original points. We can see the pixel value of red point x is the combination by nearby two blue points.



(c) Example of one dimension down sampling. The $n = 6$, $k = 2$, the distance between interpolation points is $\frac{2}{5}$. Red points are interpolation points, and blue points are original points. We can see the pixel value of red point x is the combination by nearby two blue points.

Fig. 3. Relation between interpolation points and original points.

A. Down-up Sampling Self Quotient Image (DUSSQI)

In this subsection, we demonstrate the illumination invariant preprocessing on human face images. Bilinear spline interpolation is often used in image scaling, and it only requires 4 points of original image to generate 1 point of new scaling images. As shown in Fig. 2, if we want to find the new interpolation pixel value P of the at the point (x, y), we have to compute it according to the block ratio between P and other 4 points, $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, and $Q_{22} = (x_2, y_2)$. The relation between P and Q:

$$\begin{aligned} P(x, y) = \frac{1}{(x_2 - x_1)(y_2 - y_1)} * \\ Q_{22}(x_2 - x)(y_2 - y) + Q_{21}(x - x_1)(y_2 - y) + * \\ (Q_{12}(x_2 - x)(y - y_1) + Q_{11}(x - x_1)(y - y_1)). \end{aligned} \tag{1}$$

Bilinear interpolation in 2-D images can be done by interpolating at each dimension. Firstly, we have to calculate coordinates of interpolation points. Assume that image size is $n*n$, scaling factor in one dimension is k, i is index of interpolation point, then the one dimensional coordinates of interpolation point can be calculated:

$$i * \frac{n - 1}{n * k - 1}, \tag{2}$$

if $k > 1$, it's scaling up, otherwise $0 < k < 1$ means the scaling down method. Finally, the pixel value of interpolation point can be got by calculating the distance ratio:

$$x = (x - x_1) * x_1 + (x_2 - x) * x_2, \tag{3}$$

as shown in Fig. 3(a). It's 1-D interpolation, so we have to process x axis then y axis so the 2-D images can be scaling by bilinear method. There are two example for down scaling($k = \frac{1}{2}, n = 6$) and up scaling($k = 2, n = 6$) in Fig. 3(b)(c). In Fig. 3(c) we can see that 6 points in up scaling process only reference 3 points from the down scaling points, that's why the down-up scaling images only contain low frequency component.

Freedman *et al.* [5] propose a method to obtain the smooth version image by first downsampling the image then upsampling it again. The Bilinear convolution filter processes only two values at each point in one-dimension. On the other hand, the SQI [2] method needs to compute the mean of pixel intensity in each subblock and uses the mean to determine the edge boundary of the Gaussian filter. The bilinear convolution method has less computation time than the weighted Gaussian function in SQI method. Therefore, we adopt the Bilinear interpolation scaling method to obtain low frequency components of the face image. It can be expressed as:

$$I'_k = U_k(D_k(I)), \tag{4}$$

where I'_k is the smoothed version of the input image by down-up sampling with scale of k. D, U is the down, up sampling by the Bilinear interpolation scaling method.

Here we analyse the complexity of bilinear method by the mathematical operator number. If image size is $n * n$, scaling factor in 1-D is k, we can get the operator number according to the flow of bilinear method. Firstly, computing the position of interpolation points needs $2 * (n * k)$ subtraction in one axis. Secondly, calculating the pixel value at each interpolation point needs $(k * n)^2$ addition and $2 * (k * n)^2$ multiplication in one axis. Considering bilinear image interpolation method processes in x, y axes and we need smooth version image by down-up scaling image, the total operator number can be obtained $4 * (\frac{n}{k} + n) + 6 * (\frac{n}{k}^2 + n^2)$. Table. I shows the complexity of bilinear scaling and weighted gaussian

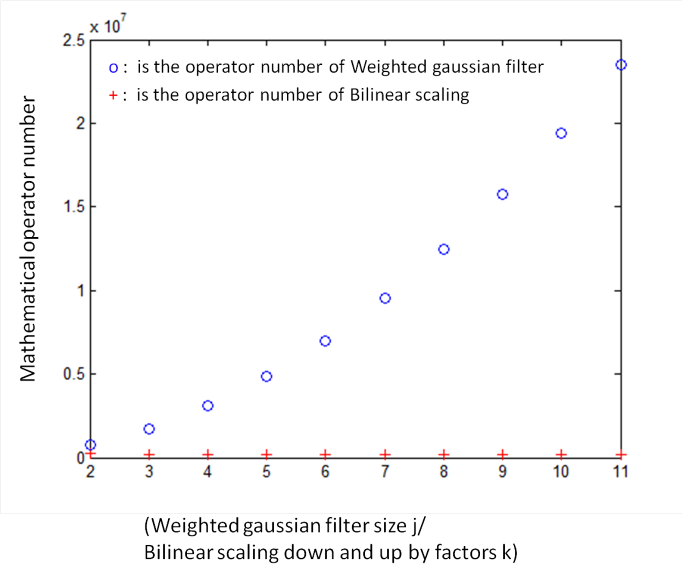


Fig. 4. Mathematical operators comparison between weighted gaussian filter and bilinear, image size is 180x180.

filter of SQI. Both them are used to generate the smooth version of input image, but take bilinear scaling as substitute for weighted gaussian filter can reduce the computational complexity dramatically. Assume the pixel number of a image is $n * n$, k is the scaling factor in one dimension, j is the pixel number of the filter in one dimension. Fig. 4 shows the comparison between these two methods. Therefore, using bilinear scaling to generate smooth image has better computational performance than weighted gaussian filter, and is more efficient to extract different scale detail from images.

TABLE I
MATHEMATICAL OPERATOR NUMBER OF BILINEAR SCALING AND WEIGHTING FILTER OF SQI.

	Mathematical operator number
SQI	$6j^2 * n^2$
Bilinear	$4 * (\frac{n}{k} + n) + 6 * ((\frac{n}{k})^2 + n^2)$

Finally, we use this low frequency component I'_k to replace the Gaussian smoothing image I' in the SQI :

$$Q_k = \frac{I}{I'_k} \tag{5}$$

where Q_k is Down-up sampling Self Quotient Image with scale k.

B. Contrast Transform and Weighting Sum

By processing DUSSQI, we can get different details from the face image. When the sampling scale of input image is large, the down-up sampling image contains more low frequency information. We can get different high-frequency information when we divide the source image with the smoothed version image to eliminate these low frequency components. Fig.5 shows the relation between the sampling scale and corresponding frequency for the DUSSQI and smoothed version images. The DUSSQI and smoothed version images contain high and low frequency information respectively. The center of the frequency domain image is the lowest frequency band. The color represents the intensity of the frequency. The intensity varies from high to low when the color changes from red to blue. We observe that the larger the down-up scaling value is, the lower frequency bands are eliminated in DUSSQI images.

In order to collect different frequency detail, we have to generate several DUSSQI images with small and large scales. Before combining these images, the image is normalize to avoid matching in different ranges. We apply a linear mapping $M(x)$ function and also truncate pixels whose luminance values are in the maximum and minimum $l\%$ range.

$$s = M(x) = \begin{cases} 1, & \textit{if } x \in \textit{max } l\% \textit{ of } X \\ 0, & \textit{if } x \in \textit{min } l\% \textit{ of } X \\ x, & \textit{otherwise} \end{cases} \tag{6}$$

where X is the DUSSQI image, and $x \in X$ is the pixel value. The l is the truncate threshold.

In this paper, we employ the combination scheme using weighted-summation rule to combine the normalized images of different DUSSQI. The weighted summation combination has an advantage in that it does not require any training phase in advance. The weighted-summation combination is expressed as:

$$S = \sum_{i=1}^{N} w_n S_n \tag{7}$$

where S_n denotes the $n - th$ normalized DUSSQI with different scale value, and S is the combination of DUSSQI with all scale values. w_n is the weighting coefficient for the $n - th$ normalized DUSSQI, such that $\sum_n w_n = 1$. Here we set $w = 1/n$ for each image.

C. Feature Extraction and Similarity Measure

McLauhlin *et al.* [4] combine the SQI method with illumination invariant feature, the Fourier transform magnitude for recognition. On the assumption that illumination influence in small block is the same, we could apply 2D Fourier magnitude as the illumination invariant feature. Each facial block is formed by their 2D Fourier magnitude which omits the phase information and allows us to take advantage of the shift invariance of the 2D Fourier magnitude representation. These presentations also tolerate little misalignment and various lighting conditions. We ignore the 0th Fourier coefficient I(0, 0), which is equivalent to eliminate the DC illumination influence. The equation can be expressed as:

$$I_{fnor}(u, v) = I_f(u, v) - I_f(0, 0) \tag{8}$$

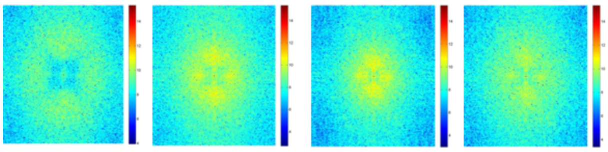
where $I_f(u, v)$ is the Fourier transform magnitude of a weighted sum DUSSQI image $S(x, y)$ and $I_{fnor}(u, v)$ is the Fourier transform magnitude of $I_f(u, v)$ whose influence of the DC illumination component $I_f(0, 0)$ is eliminated.

Here we adopt the cosine distance that is invariant to the constant multiplier. We concatenate the 2D Fourier magnitude value of each block to generate the feature vector. The cosine distance $C(a, b)$ between two vectors $a = [a_1, a_2, ..., a_n]$ and $b = [b_1, b_2, ..., b_n]$, can be expressed as:

$$C(a, b) = \frac{a \cdot b}{\|a\| \, \|b\|} \tag{9}$$

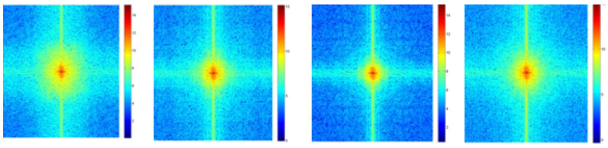
where $C(a, b)$ is the cosine angle between vector a and b. By comparing each block's 2D Fourier magnitude vector, we can determine the similarity between the gallery image and the probe image.

(a) DUSSQI



(b) Frequency domain of the DUSSQI images

(c) Down-up images(smoothed version)



(d) Frequency domain of the down-up images(smoothed version)

Fig. 5. From left to right: (a) the DUSSQI images of scaling factor of 2, 4, 6 and 2+4+6 weighted sum of these three images. (b) the frequency domain of the DUSSQI images of scaling factor of 2, 4, 6 and 2+4+6.(c) the down-up images of scaling factor of 2, 4, 6 and 2+4+6. (d) the frequency domain of the down-up images of scaling factor of 2, 4, 6 and 2+4+6.

III. Experimental Results

We describe the experimental results in this chapter. The FERET [6] and Extend YaleB database [7] are used. The computer equipment used in the experiment are: CPU(Intel Core i3 CPU 2.93GHz), 4GB RAM, 64bit OS as the computation platform, and Matlab is the implement language.

A. Extended YaleB Database

The Extended Yale B database contains facial images from 38 subjects under 64 different illumination conditions and have 9 different poses. In the experiment, we use only the frontal pose 0 in all experiments. All images are cropped to 168×192 pixels and aligned by the positions of the pupils. According to the angle between lighting and camera position, the face images are divided into 5 subsets: subset 1 ($0° \sim 12°$), subset 2 ($13° \sim 25°$), subset3 ($26° \sim 50°$), subset 4 ($51° \sim 77°$), and subset 5 ($78° \sim 90°$). We choose the image with the illumination condition of P00A+000E+00 as the gallery image for each object, and all other images are used for testing. Every face image is divided into several small blocks, and each block size is set up as 16×14 pixels. This setting reduces the misalignment between the face images. Assume that face image illumination influence on the each block is the same. It is proper to eliminate illumination influence in the block by removing the (0,0) coefficient of Fourier magnitude. The linear mapping $M(x)$ function used to truncate pixels are in the maximum and minimum 2% range.

Different illumination invariant preprocessing were used here (Figure. 6) and Fourier magnitude(FM) was the recognition feature. Block size was set as 16x14 and cosine distance is the similarity comparison method. SQI+FM [8] use the self quotient image(filter size are 3x3 and 9x9) and Fourier magnitude. Tan method [9] use gamma correction($\gamma = 0.2$.) and different of gaussian($\sigma = 1, 2$.). DCT-nor [10] eliminate illumination influence by setting DCT coefficients to zero(here set 50 coefficients to zero.). Gradient [11] computes the orientation of the image gradients in each pixel of the face images and uses the computed face representation as an illumination invariant version of the input image. Single scale Retinex(SSR) and multi scale Retinex(MSR) [12] based on the Retinex theory, a face image is decomposed into its illumination invariant part and its smoothed version which can be obtained from convolution an original image by Gaussian filter(SSR filter size is 15x15, MSR filter size are 7x7, 15x15, 21x21). Single scale self quotient image(SSSQI) and multi scale self quotient image(MSSQI) [1] also based on the Retinex theory, but SQI use weighted gaussian filter(SSSQI filter size is 5x5, $\sigma = 1$. MSSQI filter size are 3x3, 7x7, $\sigma = 1, 1.1$.) to generate smooth version images.

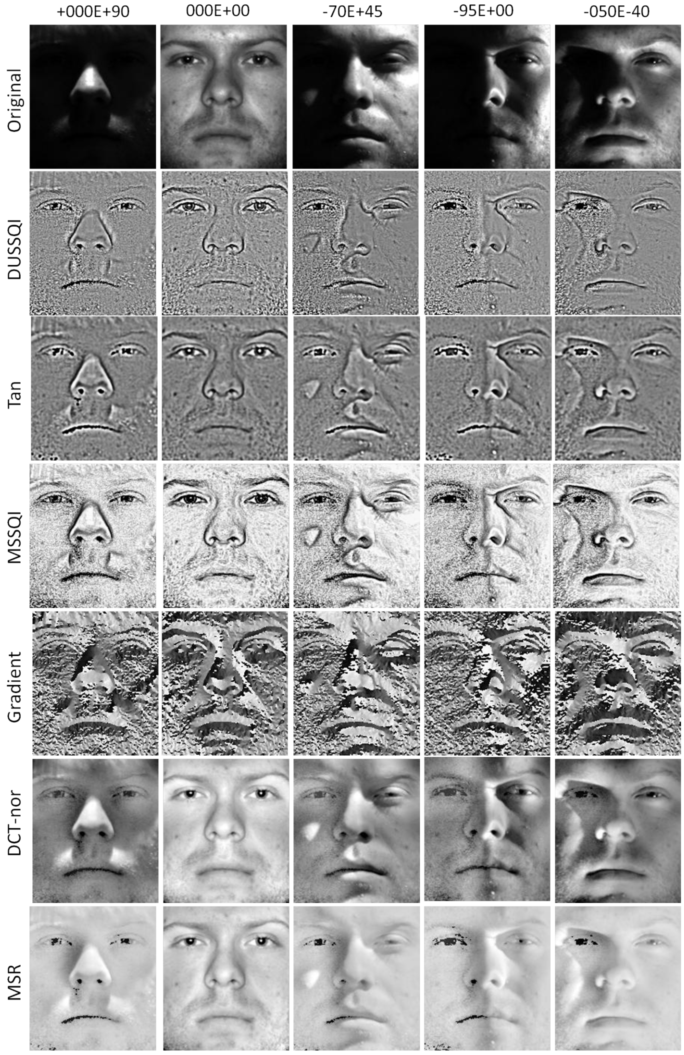


Fig. 6. Different preprocessing methods are used in extended yaleb object1. Each column is different lighting conditions, and each row is different preprocessing methods.

In the first experiment, Fig. 7 and Table. II show the recognition rate comparison for our approach down-up sampling self quotient image(DUSSQI) with other illumination invariant face recognition methods. We can see that SQI [8], SSSQI, MSSQI [1] have high recognition rate but the computation time are much more than DUSSQI. In fact, SQI, SSSQI, MSSQI have 35, 29.9, 33.7 times computation time than DUSSQI. Because the Bicubic convolution filter processes only 2 pixels at each point in one-dimension, and the SQI method has to compute the mean of pixel intensity in each subblock to determine the weighting of the Gaussian filter. Therefore, the DUSSQI method has higher computation performance than the weighted Gaussian filter in SQI method, and DUSSQI has better recognition rate than original weighted Gaussian filter in SQI. DCT-nor [10], gradient [11], SSR [12], MSR [13] have good computational performance, but the ability to resist illumination variation face recognition can't compete with other methods. It's worth to mention, Tan [9] use the gamma correction and different of gaussian to eliminate illumination influence, and has high computational performance, high recognition rate. However, the details that DUSSQI provided are more than Tan method, so DUSSQI has higher recognition rate than Tan method.

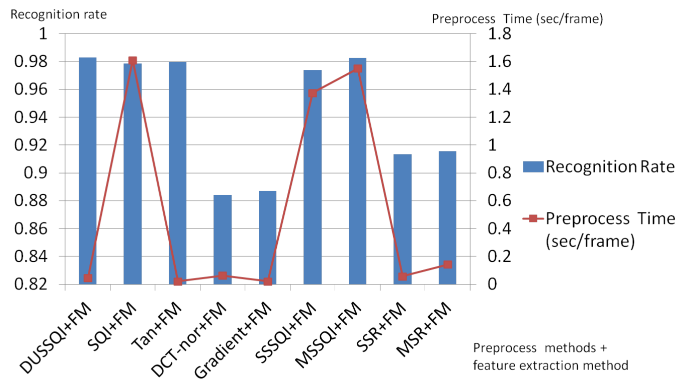


Fig. 7. Recognition rate(left axis) and preprocessing time(right axis) with different preprocessing methods in Extended Yaleb database. Red line is preprocessing time and blue rectangles are recognition rate. There are different preprocessing methods, Down-up scaling self quotient image(DUSSQI), Self quotient image(SQI) [8], Tan method [9], Discrete cosine transform normalization (DCT-nor) [10], Gradient face [11], Single scale self quotient image (SSSQI) [1], Multi scale self quotient image (MSSQI) [1], Single scale Retinex (SSR) [12], Multi scale Retinex (MSR) [13].

TABLE II
THE RECOGNITION RATE, PREPROCESSING TIME AND PREPROCESSING TIME RATIO WITH DIFFERENT PREPROCESSING METHODS IN EXTENDED YALEB DATABASE.

Preprocessing method	Recognition Rate	Preprocess Time(sec/frame)	Preprocess Time ratio
DUSSQI+FM	**0.983**	**0.0456**	**1**
SQI+ FM	0.9785	1.6098	35
Tan + FM	0.9797	0.0211	0.458
DCT-log+FM	0.884	0.0634	1.378
Gradient+FM	0.8869	0.0224	0.486
SSSQI+FM	0.9739	1.3749	29.89
MSSQI+FM	0.9826	1.551	33.72
SSR+FM	0.9134	0.0576	1.252
MSSQI+FM	0.9155	0.1425	3.098

B. FERET Database

The images in FERET contain variations in lighting, facial expression, aging etc.. In this work, only frontal faces are considered. The subset Fa containing 1,196 frontal images of 1,196 subjects was used as the gallery set, while the subset Fc (194 images taken under different illumination conditions) was used as the probe sets. We focus on illumination variation, so only subset Fc was used in our experiment. All the Images used in the experiment are cropped into 180x180.

We take three methods that have high recognition rate in the first experiment into the second experiment, the recognition rates are shown in Fig. 8 and processing times are shown in Fig. 9. Different block sizes with Fourier magnitude feature are used in this experiment. The figure shows that MSSQI and DUSSQI have the highest recognition rate when block size is 45x45 and this block size is bigger than the block size used in Extended Yaleb database. It's because that images in FERET database have more misalignment than in Extended Yaleb database, as shown in Fig. 10, we can find some misalignment and face expression change in these face images. From this experiment, the robustness to the misalignment and face expression change can be approved. Moreover, DUSSQI also has high recognition rate, high computational performance in FERET database, and just need to change block size to fit different situations.

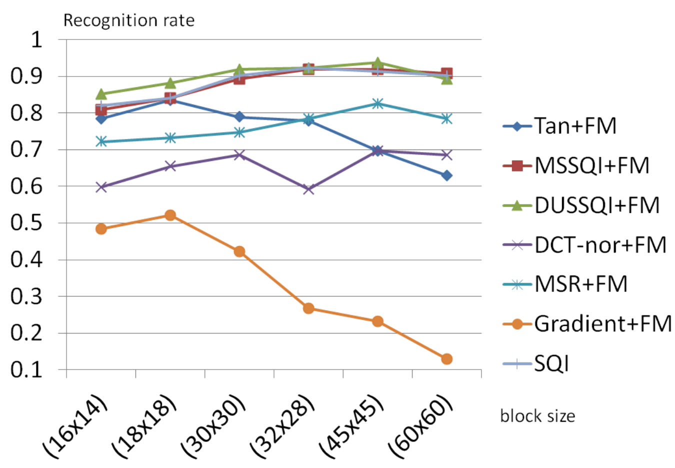


Fig. 8. Tan, DUSSQU, MSSQI, DCT-nor, MSR, Gradient, SQI are tested in FERET database with different block size.

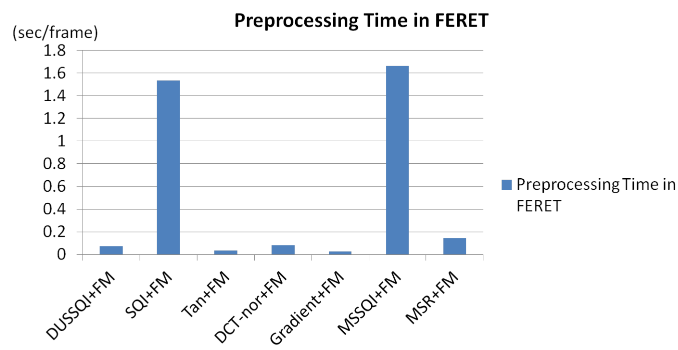


Fig. 9. The processing time with different preprocessing methods in FERET method.

Fig. 10. Images in FERET database. First column in the figure has different normal images in subset Fa, and the second column has different lighting conditions images in subset Fc.

IV. CONCLUSION

We present a new illumination invariant face recognition algorithm, Down-Up Sampling Self Quotient Image (DUSSQI), which achieves very high recognition accuracy on the extended YaleB database(98.3%), and FERET database(93.85%). The proposed algorithm has better recognition rate than the original SQI method, and reduces 97.1% computational time compared to that of SQI. The scheme we provided is more robust to real-time system. Furthermore, we presents a new idea that retrieves complete illumination invariant image by combining different frequency detail from down-up scaling images. Finally, we extract the information more specifically to get better illumination invariant image for recognition, and provided a method to handle illumination invariant face recognition problem in the fast recognition system.

REFERENCES

[1] Haitao Wang, S.Z. Li, and Yangsheng Wang, "Face recognition under varying lighting conditions using self quotient image," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, may 2004, pp. 819 – 824.

[2] Edwin H. Land, John, and J. Mccann, "Lightness and retinex theory," *Journal of the Optical Society of America*, pp. 1–11, 1971.

[3] Virendra P. Vishwakarma, Sujata Pandey, and M. N. Gupta, "An illumination invariant accurate face recognition with down scaling of dct coefficients.," *CIT*, vol. 18, no. 1, pp. 53–67, 2010.

[4] N. McLaughlin, Ji Ming, and D. Crookes, "Illumination invariant facial recognition using a piecewise-constant lighting model," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 1537 –1540.

[5] Gilad Freedman and Raanan Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, pp. 12:1–12:11, Apr. 2011.

[6] P.J. Phillips, Hyeonjoon Moon, S.A. Rizvi, and P.J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[7] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[8] N. McLaughlin, Ji Ming, and D. Crookes, "Illumination invariant facial recognition using a piecewise-constant lighting model," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 1537 –1540.

[9] Xiaoyang Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[10] W. Chen, Meng-Joo Er, and Shiqian Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, no. 2, pp. 458–466, 2006.

[11] Taiping Zhang, Yuan-Yan Tang, Bin Fang, Zhaowei Shang, and Xiaoyu Liu, "Face recognition under varying illumination using gradientfaces," *Image Processing, IEEE Transactions on*, vol. 18, no. 11, pp. 2599–2606, 2009.

[12] D.J. Jobson, Z.-u. Rahman, and G.A. Woodell, "Properties and performance of a center/surround retinex," *Image Processing, IEEE Transactions on*, vol. 6, no. 3, pp. 451–462, 1997.

[13] 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, pp. 965–976, 1997.