Weighted Feature-Level Fusion of Color Local Texture Features for Face Recognition

Thanh-Dung Dang and Xuan-Thu Tuong Thi

***Abstract*—This paper presents an extension of the framework proposed by Choi *et al*. that uses color local texture features (CLTF) for face recognition. The original framework combines multiple CLTFs, each of which corresponds to the associated color channel, with no weights at feature-level. In the proposed extension, the combination is performed with weights that are calculated based on the contribution of each channel to recognition performance. After the combination, PCA is used to select in the combining result the most important components which are used as the unique feature vector representing a face image. Comparative experiments have been conducted to evaluate the customized framework for FR on three public face databases, i.e., CMU-PIE, Color FERET and Postech Face's 01. Experimental results show that the customized framework yields better recognition rates than the framework by Choi *et al*.**

***Index Terms*—Color face recognition (FR), color local texture features, color spaces, feature combination, local binary pattern (LBP).**

I. Introduction

Face recognition (FR) has a wide range of applications, such as biometric identification, video surveillance, face indexing in multimedia contents etc. Thus it has been receiving an increasing interest in research on computer vision and pattern recognition. An important step of the FR process is to extract features representing for a face image. Among many proposals for extracting features, local texture features [1] have been considered as powerful face descriptor. This is because it is evident that local texture features are more robust to variations of facial expression, pose, and occlusion. Particularly, local binary pattern (LBP) texture features [2] and Gabor wavelets texture features [3] provide highly discriminative information for FR.

More and more research on textured image analysis has been focusing on the color aspects of the analysis [4]. These works empirically show that color information keep a complementary role in classification/recognition and texture analysis. Thus, it can be used to enhance classification/recognition performance.

Mäenpää & Pietikäinen [4] empirically compared the effectiveness of three methods of extracting features for classification tasks, including grayscale texture, color indexing, and color texture. The experiment was carried out on a data set of texture images taken under either a varying or static (constant) condition of illumination. The results from this experiment indicate that, under the static illumination, condition, color texture features generally outperform grayscale texture feature.

Drimbarean & Whelan [5] investigated the extensions of three grayscale texture methods, including Gabor filtering, cooccurrence, and linear transform, into color images. The found that color information helps improve classification performance.

The work by Paschos [6] indicates that incorporating color information into a texture analysis is beneficial for recognition/classification tasks. Moreover, the author showed that color texture analysis in perceptually uniform color spaces such as HSV (Hue, Saturation, and Value) and $L \times a \times b$ yields better results than in RGB (Red, Green, and Blue) color space.

In summary, the results from the aforementioned studies suggest that FR performance when using the combination of color and texture information is better than FR performance when using only color or texture information. The remaining question needs to be answered is how to combine these information such that both color and texture information effectively contribute to FR performance. Choi, Ro, & Plataniotis [7] proposed a framework in which color and texture information are effectively combined to improve FR performance. The experimental results from their research show that their proposed color local texture features (CLTFs) help improve significant recognition rates for face images taken under severe variation in illumination. Besides, their approach also yields better recognition rates than FR approaches using only color or texture information.

Despite providing a good improvement in FR, the author pointed out several limitations in their framework. One of the limitations of their framework is to treat color local texture features from different channel equally when combining these features to form a unique feature representing a facial image. This is improper because it has been observed in [8], [9] that local texture feature from different channels contributes differently to FR performance. Thus, in this paper, we propose a more effective weighted feature-level fusion scheme in which color local texture features are combined with different weights. The weight of each color local texture feature is determined based on the extent of the contribution of that feature to FR performance. This contribution can be identified by learning from a small training set of face images.

II. The Extended Framework of FR Using Color Local Texture Features

As shown in Fig. 1, the extended framework for FR using

Manuscript received October 7, 2014; revised December 10, 2014.

Thanh-Dung Dang is with the Faculty of Information Technology HCMC University of Technical Education Ho Chi Minh City, Vietnam (e-mail: dungdt@hcmute.edu.vn).

Xuan-Thu Tuong Thi is with Faculty of Information Technology HCMC University of Foreign Languages – Information Technology Ho Chi Minh City, Vietnam (e-mail: tuongxuanthu@huflit.edu.vn).

color local texture features consists of five major steps: color space conversion, partition and feature extraction, weighted combination, PCA, and classification.

The first step is to align color face images in RGB by translating and rescaling to a fixed template. Subsequently, color space conversion is performed to convert an image in RGB into a new image represented in another color space. Then, each of the spectral images of the face image in the new color space is partitioned into local regions.

After that, texture feature for each of the local regions is separately and independently computed. These texture features are called “color local texture features” because they are extracted from the local face regions that are obtained from different color channels (spectral images).

Of note, the key point in the method of FR using color information is to extract the opponent texture features [10] from each pair of spectral images, and to extract unichrome (or channelwise) texture features from each individual spectral image. The combination of these two kinds of features allows for improving FR performance, as compared with grayscale texture feature that is extracted based on only the luminance of an image.

Since many CLTFs are available (each of these features is computed for a spectral image), we need to combine them in order to form a unique feature representing a color face image. This unique feature then is classified using a particular classifier. In the original framework by Choi *et al.* [7], this combination is a non-weighted combination. However, in this paper, we propose to use weighted combination, as described in Section IV. After the weighted combination, PCA is used to select the most important information from the CLTFs for classification.

III. Extraction of Color Local Texture Features for Each Spectral Image

The CLTFs used in our framework are color local binary patterns (LBP) that are an extension of grayscale LBP proposed by [11].

Let S_i(i=1, …, K) be K different spectral (color-component) images of a color face image, the unichrome (or channelwise) LBP feature for each channel is independently and separately calculated from each S_i. This unichrome LBP feature is computed using the uniform.

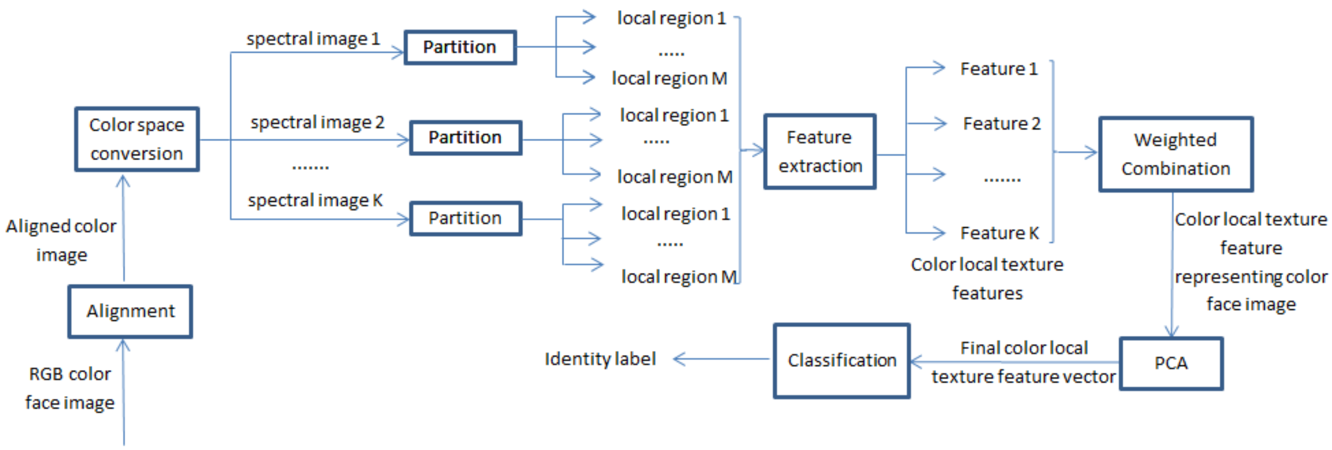


Fig. 1. Extended FR framework based on color local texture feature.

LBP operator in [2]. The reason of adopting this uniform LBP operator is because, as reported in [2], a typical face image contains only a small number of uniform patterns. Of note, a uniform pattern is a LBP value whose binary representation contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular [2].

Let z_c denotes the center pixel position of S_i, and z_n ($n=0, \ldots, P-1$) denote spaced pixels (or sampling points) on a circle of radius that form a circular neighborhood of the center pixel z_c. Then, the unichrome LBP operation for the center pixel position z_c of S_i is defined as follows [2]:

$$LBP_{P,R}^{(i)}(z_c) = \begin{cases} \sum_{n=0}^{P-1} \delta(r_n^{(i)} - r_c^{(i)})2^n \ if\ H \le 2 \\ P(P-1)+2 \quad otherwise \end{cases} \quad (1)$$

where

$$H = \left|\delta\left(r_{P-1}^{(i)} - r_c^{(i)}\right) - \delta\left(r_0^{(i)} - r_c^{(i)}\right)\right| + \sum_{n=1}^{P-1}\left|\delta\left(r_{P-1}^{(i)} - r_c^{(i)}\right) - \delta\left(r_{n-1}^{(i)} - r_c^{(i)}\right)\right| \quad (2)$$

and

$$\delta(x) = \begin{cases} 1, x \ge 0 \\ 0, x < 0 \end{cases}$$

$r_n^{(i)}(n = 0, \ldots, P-1)$ denotes the pixel values of S_i at z_c, and $r_c^{(i)}$ denotes the pixel value at z_c of a circular neighborhood.

Now, let L_i be an LBP image corresponding to S_i. An LBP image is formed as following: Each of the pixel values of $L_i(x, y)$ is filled with $LBP_{P,R}^{(i)}(z_c)$ at its given pixel location $z_c=(x, y)$. In the case of uniform pattern, that is, when number of transitions is smaller or equals to 2 (i.e., $H \le 2$), each pixel of $L_i(x, y)$ is labeled as a value of $\sum_{n=0}^{P-1} \delta(r_n - r_c)2^n$. Otherwise, it will be labeled as constant $P(P$-1)+2.

Of note, the aforementioned unichrome LBP operation is performed on each spectral image of a color face image.

In order to encode and reflect the local properties of a spectral image S_i, a regional LBP pattern histogram for each local region $S_i^{(m)}$ is computed. A regional LBP pattern histogram consists of $P(P\text{-}1)+3$ bins. These bins include $P(P\text{-}1)$ bins for the patterns with two transitions, 2 bins for the patterns with zero transitions, and 1 bin for all non-uniform patterns.

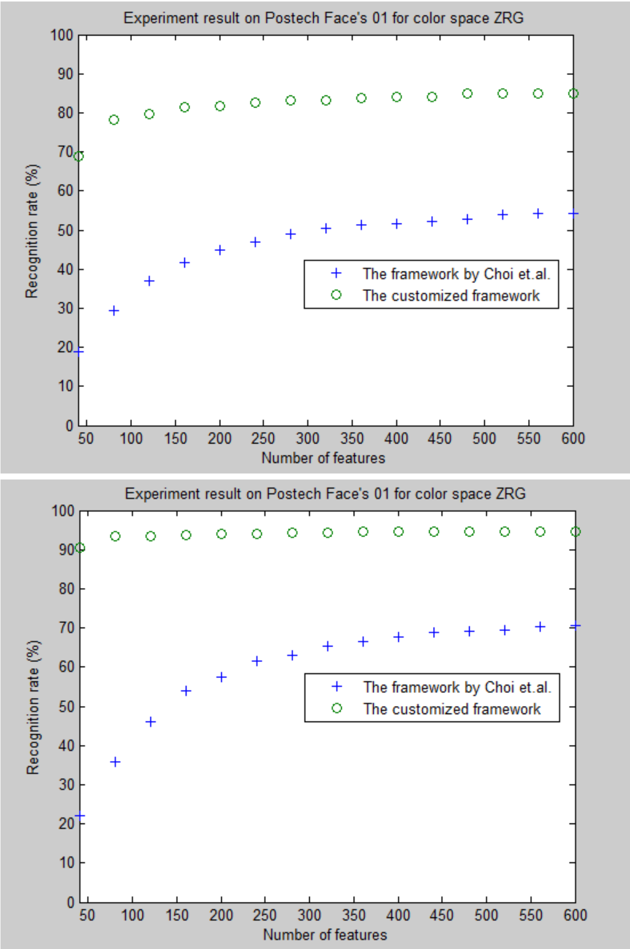


Fig. 2. Rank-1 identification rates for dataset Postech Face's 01 when using color space ZRG.

The LBP histogram for the m^{th} local region $S_i^{(m)}$ is computed as follows:

$$h_i^{(m)}(k) = \sum_{z_c=(x,y)\in S_i^{(m)}} T\left(L_i\left(z_c\right)=k\right)$$

$$for\ 1 \le k \le P\left(P-1\right)+3 \qquad (3)$$

where

$$T(A) = \begin{cases} 1, & if\ A\ is\ true \\ 0, & if\ A\ is\ false \end{cases} \qquad (4)$$

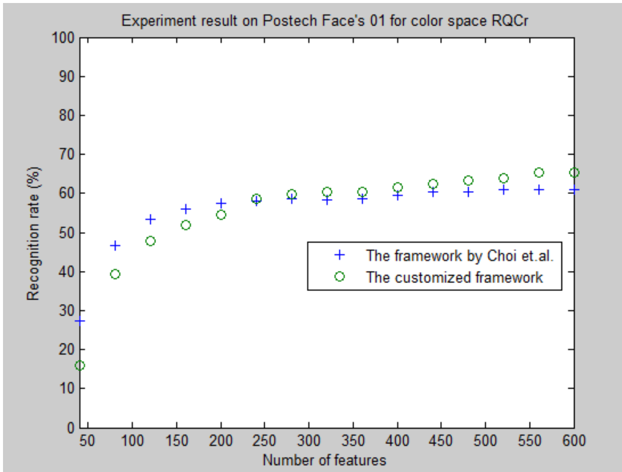
and k denotes the k^{th} LBP value in the range of [0, $P(P\text{-}1)+2$]. Thus, $h_i^{(m)}(k)$ is the number of pixels (in the LBP image) whose values equal to the k^{th} LBP value in the range of [0, $P(P\text{-}1)+2$].

The regional unichrome LBP descriptor for $S_i^{(m)}$ is defined as following:

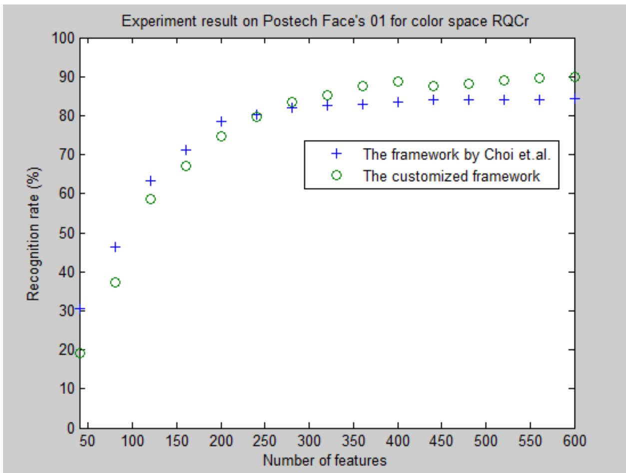
$$h_i^{(m)} = \left[h_i^{(m)}(1), h_i^{(m)}(2), \ldots, h_i^{(m)}(P(P-1)+3)\right]^T \quad (5)$$

Of note, $h_i^{(m)}$ computed through (3) encodes regional LBP histogram information for only one local region $S_i^{(m)}$. In order to encode the information about the spatial relation of these facial local regions, all of the $h_i^{(m)}$ $(m=1,\ldots,M)$ values are concatenated into a single column vector. This vector is called unichrome LBP feature for the spectral image S_i. The unichrome LBP feature thus can be mathematically expressed as follows:

$$x_i^{U\_LBP} = \left[\left(h_i^{(1)}\right)^T \left(h_i^{(2)}\right)^T \ldots \left(h_i^{(M)}\right)^T\right]^T \qquad (6)$$



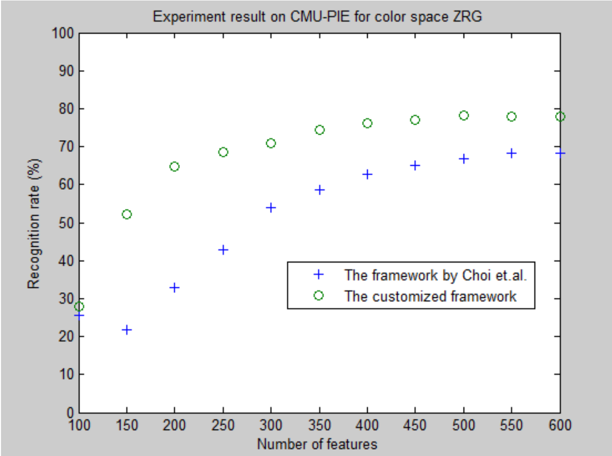
size(G) = 5, size(G_{LW}) = 2



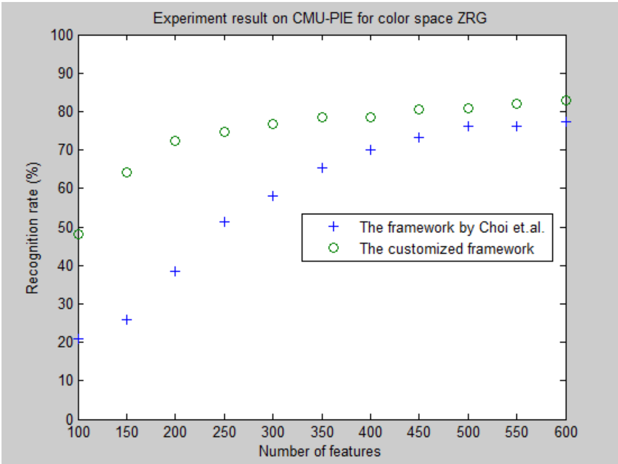
size(G) = 8, size(G_{LW}) = 3

Fig. 3. Rank-1 identification rates for dataset Postech Face's 01 when using color space RQCr.

Beside the unichrome LBP feature, the opponent LBP features are also computed. An opponent LBP feature is calculated for each pair of different color channel (spectral image). To this end, the center coordinate for a neighborhood and the neighbor itself are taken from two different color channels.



size(G) = 8, size(GLW) = 4



size(G) = 10, size(GLW) = 4

Fig. 4. Rank-1 identification rates for dataset CMU-PIE when using color space ZRG.

The opponent LBP value for a pair of spectral images S_i and S_j ($i \neq j$) for the center pixel z_c in S_i is computed as follows:

$$LBP_{P,R}^{(i)}(z_c) = \begin{cases} \sum_{n=0}^{P-1} \delta(r_n^{(i)} - r_c^{(i)}) 2^n \; if\; H \leq 2 \\ P(P-1)+2 \quad otherwise \end{cases} \tag{7}$$

(for $i \neq j$)

where

$$H = \left|\delta\left(r_{P-1}^{(j)} - r_c^{(i)}\right) - \delta\left(r_0^{(j)} - r_c^{(i)}\right)\right| + \sum_{n=1}^{P-1} \left|\delta\left(r_n^{(j)} - r_c^{(i)}\right) - \delta\left(r_{n-1}^{(j)} - r_c^{(i)}\right)\right| \tag{8}$$

and $r_n^{(j)}$ $(n = 0, \ldots, P-1)$ denotes the pixels values of S_j that form a circular neighborhood of the center pixel z_c of S_i, and $r_c^{(i)}$ denotes a pixel value of S_i at its coordinate z_c.

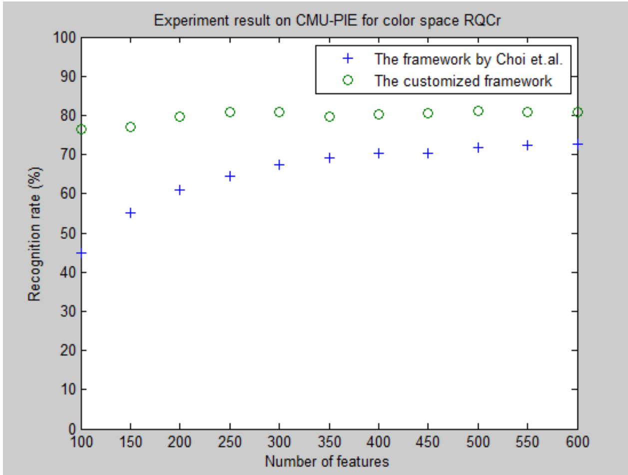
The opponent LBP feature for the pair of S_i and S_j, denoted by $x_{i,j}^{O\_LBP}$, is computed by using (3), (5), (6), and (7).

The color LBP (CLBP) feature representing for a spectral image is defined as following:

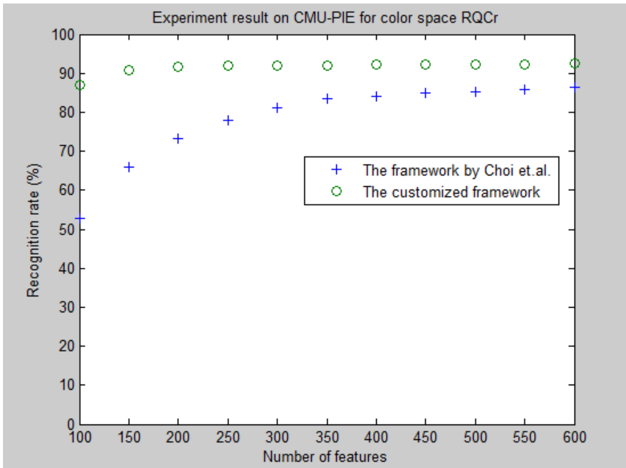
$$x_i^{CLBP} = \left[\left(x_i^{U_{LBP}}\right)^T \left(x_{i,j}^{O_{LBP}}\right)^T \ldots \left(x_{i,K}^{O_{LBP}}\right)^T\right]^T \tag{9}$$

(for $i \neq j$)

Of note, the CLBP feature for the i^{th} spectral image, x_i^{CLBP}, is the concatenation of 1 unichrome LBP feature and $K-1$ opponent LBP features computed from S_i and $K-1$ different spectral images S_j ($j = 1, \ldots, K$ and $i \neq j$). The CLBP feature for a spectral image calculated by (9) contains more discriminating information than what a single grayscale LBP operation can provide [7].



size(G) = 8, size(GLW) = 4



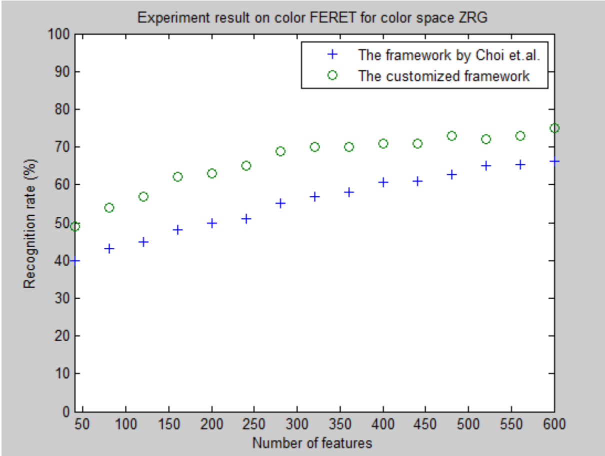
size(G) = 10, size(GLW) = 4

Fig. 5. Rank-1 identification rates for dataset CMU-PIE when using color space RQCr.

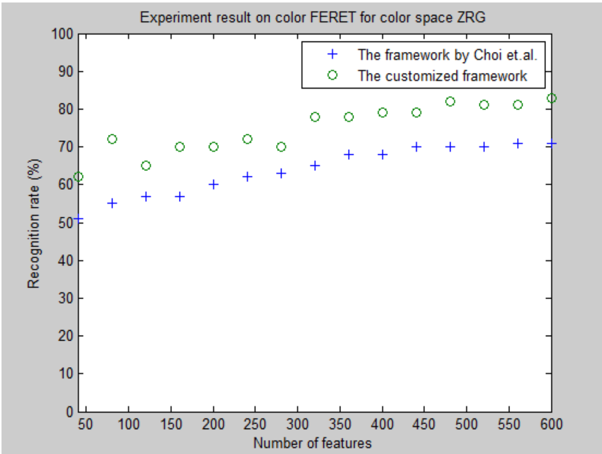
IV. Weighted Combination of CLTFs from Individual Channels

Let x^i denote a CLTF obtained from a spectral image S_i. Besides, let I^P be an unknown color face image to be identified, which is denoted as a probe. In addition, let G be a gallery set consisting of prototype enrolled color face images, each of which is denoted by I^G, of known individuals (i.e., $I^G \in G$). Furthermore, without any loss of generality, we denote the individual CLTFs of I^P and I^G by x_i^P and x_i^G, respectively, where $i = 1, \ldots, K$.

In the original framework by Choi *et al.*, before the non-weighted combination of CLTFs from individual spectral images, a low-dimensional feature extraction technique is employed on each of the CLTFs to avoid degradation in the FR performance caused by the high dimensionality and the redundant information.



size(G) = 5, size(GLW) = 2



size(G) = 8, size(GLW) = 3

Fig. 6. Rank-1 identification rates for dataset color FERET when using color space ZRG.

Let us denote the i^{th} face feature extractor (e.g., PCA) by φ_i for extracting a low-dimensional feature of the CLTF x_i. Note that φ_i can be formed with a training set of CLTFs X_i, all of which are computed from a set of spectral training images. Then, the low-dimensional features of I^P and I^G are obtained as follows (using the corresponding φ_i):

$$f_i^P = \varphi_i\left(x_i^P\right) \qquad f_i^G = \varphi_i\left(x_i^G\right) \tag{10}$$

where $i = 1, \ldots, K$.

Then, K complementary low-dimensional features, given by (10), are combined at the level of features (by concatenating low-dimensional features in the column order), i.e.,

$$f^P = \left[\left(w_1 f_1^P\right)^T \ldots \left(w_K f_K^P\right)^T\right]$$
$$f^G = \left[\left(w_1 f_1^G\right)^T \ldots \left(w_K f_K^G\right)^T\right] \tag{11}$$

where $0 \le w_i \le 1,\ i = 1, \ldots, K$ are weights associated with each channel.

Let G_{LW} be a set of face images used for learning weights. Denote $P_{LW} = G - G_{LW}$, then the weights are identified by the following procedure:

1) From the set G_{LW}, identify the low-dimensional feature extractor φ_i^{LW} (e.g., PCA).

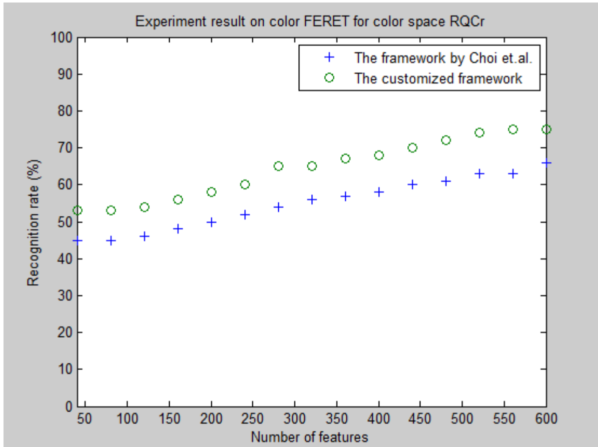
2) The weights are identified by:

$$\left(w_1, \ldots, w_K\right) = \arg\max_{0 \le w_i \le 1}(N) \tag{12}$$

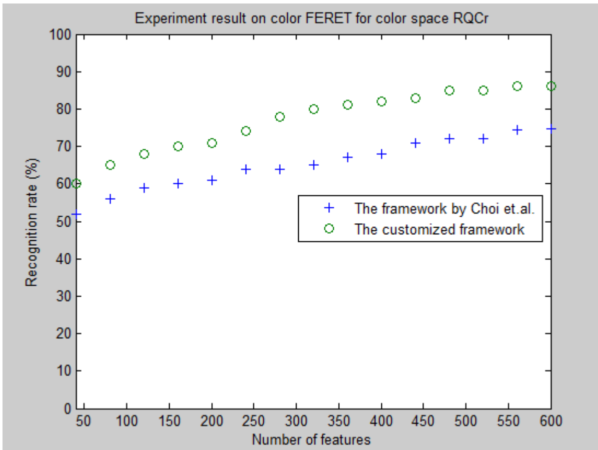
where N is the number of correct classification of $f^{P_{LW}}$.

$$f_i^{P_{LW}} = \varphi_i^{LW}(x_i^{P_{LW}}),\ i = 1, \ldots, K \tag{13}$$

$$f^{P_{LW}} = \left[\left(w_1 f_1^{P_{LW}}\right)^T \ldots \left(w_K f_K^{P_{LW}}\right)^T\right] \tag{14}$$



size(G) = 5, size(GLW) = 2



size(G) = 8, size(GLW) = 3

Fig. 7. Rank-1 identification rates for dataset color FERET when using color space RQCr.

V. Reducing Dimensions of the Vector Resulting from Weighted Combination

Before the weighted combination, each of the CLTFs from the individual channels has been applied PCA in order to reduce its number of dimensions to d. This means that each CLTF contributes d components to the combining result. However, because information from each channel can contribute differently to discrimination, the channel that plays a more important role to discrimination should contribute larger number of components of its CLTF to the final feature vector used for classification. Thus, after the weighted combination, we apply PCA one more time on the result of the combination to select the most important components among those from the K CLTFs. By this way, a channel with more important information to discrimination will contribute more components to the final feature vector, which is only the representation of the face image.

VI. Experimental Results

Three publicly available face dataset, i.e., CMU-PIE [12], Postech Faces' 01 [13], and Color FERET [14] were used to evaluate the feature extracted by the customized framework. For the CMU-PIE dataset, 1428 frontal images of 68 subjects (21 images of different illumination variations per subject) were used. The Postech Faces' 01 dataset contains 1819 true-color face images of 107 Asians (each person has 17 images taken under various conditions: 1 normal face, 4 illumination variations, 8 pose variations, 4 expression variations). From the Color FERET dataset, 1378 face images of 107 subjects were selected for the experiments. Besides, only the rotated face images that both eyes can be reliably identified for normalization were collected. These images included five different pose angles ranging from -45^o to 45^o and had neutral expression and illumination.

The Mahalanobis distance [15] was used for the NN classifiers. LBP parameters chosen included *P*=8 and *R*=2. As explained in [7], two color spaces, *ZRG* and *RQCr*, were used to compare FR accuracy when using the original and customized frameworks. Before applying the frameworks, the facial images were aligned in the same way used by Choi *et al*. [7].

Each of the three datasets was randomly divided into the training and probe sets as described in Table I.

Figures from Fig. 2 to Fig. 7 show that the customized framework outperforms the original framework by Choi *et al*. Furthermore, the difference between the two frameworks is more significant when the number of features increases. However, when the number of features is around 600, the difference becomes stable.

TABLE I: Number of Subjects in G and G_{LW}

	ZRG	ZRG	RQCr	RQCr
Postech Face's 01	G = 5 G_{LW} = 2	G = 8 G_{LW} = 3	G = 5 G_{LW} = 2	G = 8 G_{LW} = 3
CMU-PIE	G = 8 G_{LW} = 4	G = 10 G_{LW} = 4	G = 8 G_{LW} = 4	G = 10 G_{LW} = 4
Color FERET	G = 5 G_{LW} = 2	G = 8 G_{LW} = 3	G = 5 G_{LW} = 2	G = 8 G_{LW} = 3

VII. Conclusion

This paper has investigated the contribution of weighted feature fusion for improving the FR performance under the Choi et al. framework. The proposed fusion scheme includes two steps: 1) to perform weighted feature fusion using the weights identified by the method mentioned in Section IV, and 2) to apply PCA to the combining result. Experimental results in this research reveal that the proposed scheme significantly outperforms the original fusion method by Choi *et al*. Furthermore, this scheme considerably improves the FR accuracy when recognizing face images taken under a severe change in illumination, as compared with their no-weight fusion counterpart. More importantly, among the three datasets were tested in this research, the FR accuracy is enhanced most in the case of the Postech dataset, in which face images were taken under variations of illumination, pose, and expression. This means that the proposed fusion scheme effectively exploits the discriminating information in individual channels to make the combining result be more robust to variations of pose, expression, and illumination.

References

[1] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2617-2628, Oct. 2007.

[2] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.

[3] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467-476, Apr. 2002.

[4] T. Mäenpää and M. Pietikäinen, "Classification with color and texture: jointly or separately?" *Pattern Recognition*, vol. 37, no. 8, pp. 1629-1640, Aug. 2004.

[5] A. Drimbarean and P. Whelan, "Experiments in colour texture analysis," *Pattern Recognition Letters*, vol. 22, no. 10, pp. 1161-1167, 2001.

[6] G. Paschos, "Perceptually uniform color spaces for color texture analysis: An empirical evaluation," *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 932-937, Jun. 2001.

[7] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color local texture features for color face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1292–1309, Oct. 2012.

[8] H. Stokman and T. Gevers, "Selection and fusion of color models for image feature detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 371–381, Mar. 2007.

[9] P. Shih and C. Liu, "Comparative assessment of content-based face image retrieval in different color spaces," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 7, pp. 873-893, Nov. 2005.

[10] A. Jain and G. Healey, "A multiscale representation including opponent color features for texture recognition," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 124–128, Jan. 1998.

[11] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proc. the 12th IAPR International Conference on Pattern Recognition*, pp. 582-585, 1994.

[12] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[13] H. S. Lee, S. Park, B. N. Kang, J. Shin, J. Y. Lee, H. Je, and D. Kim, "The POSTECH face database (PF07) and performance evaluation," *In FG'2008. 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1-6, 2008.

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

[15] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recognition Letters*, vol. 25, no. 6, pp. 711–724, Apr. 2004.

Thanh-Dung Dang was born in Vietnam in 1972. He received his Ph.D. degree from the Department of Computer Science, National Taiwan University, Taiwan in 2012. He is currently a lecturer with Faculty of Information Technology, Ho Chi Minh City University of Technology and Education, Vietnam.

Xuan-Thu Tuong Thi was born in Vietnam in 1984. She received her M.S. degree from the Department of Computer Science, Chung-Yuan Christian University, Taiwan in 2012. She is currently a lecturer with Faculty of Information Technology, Ho Chi Minh City University of Foreign Language–Information Technology, Vietnam.