Accepted Manuscript for Publication in IEEE Transaction on Image Processing

Boosting Color Feature Selection for Color Face Recognition

Jae Young Choi, *Student Member*, *IEEE*, Yong Man Ro, *Senior Member, IEEE,* and

Konstantinos N. Plataniotis, *Senior Member, IEEE*

Abstract

This paper introduces the new color face recognition (FR) method that makes effective use of boosting learning as color-component feature selection framework. The proposed boosting color-component feature selection framework is designed for finding the best set of color-component features from various color spaces (or models), aiming to achieve the best FR performance for a given FR task. In addition, to facilitate the complementary effect of the selected color-component features for the purpose of color FR, they are combined using the proposed weighted feature fusion scheme. The effectiveness of our color FR method has been successfully evaluated on the following five public face databases (DBs): CMU-PIE, Color FERET, XM2VTSDB, SCface, and FRGC 2.0. Experimental results show that the results of the proposed method are impressively better than the results of other state-of-the-art color FR methods over different FR challenges including highly uncontrolled illumination, moderate pose variation, and small resolution face images.

Index Terms

Color face recognition, boosting learning, color space, color-component, feature selection, weighted feature fusion.

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

1. Introduction

Recently, considerable research work in face recognition (FR) has shown that facial color information can be used to considerably improve FR performance, compared to the FR methods relying only on grayscale information [1-7]. In particular, it has been reported in [5-6] that the effectiveness of color information can become significant for improving FR performance when face images are taken under strong variations in illumination, as well as with low spatial resolutions.

In general, the three components of a color can be defined in many different ways leading to a wide variety of color spaces [8]. In addition, it has been observed in [8-10] that different color spaces (or color models) possess distinct characteristics and effectiveness in terms of discriminating power for visual classification task. This suggests that different color components can often provide different complementary information to the specific classification task [9-10]. Hence, an optimal subset of color components may not be unique for different classification or pattern recognition problems.

Most of the existing color FR methods (including our previous work [5]) are restricted to using a fixed color-component configuration comprising of only "two" or "three" color components (like YQC_r [1] from YC_bC_r and YIQ color spaces). In particular, currently used color-component choices are mostly made through a combination of intuition and empirical comparison [1-3], [5-7] without any systematic selection strategy. As such, existing methods may have a limitation to attaining the best FR result for given FR task. This is because specific color components effective for a particular FR problem could not work well for other FR problems under other FR operating conditions (e.g., illumination variations) that differ from those considered during the process of determining specific color components. Hence, the important issue in color FR is: how can one select the color components from various color models in order to achieve the best FR performance for the specific FR task?

In this paper, to cope with the aforementioned issue, we propose a new color FR method. Our method takes advantage of "boosting" learning [13] as a feature selection mechanism, aiming to find the optimal set of color-component features for the purpose of achieving the best FR result. To the best of our
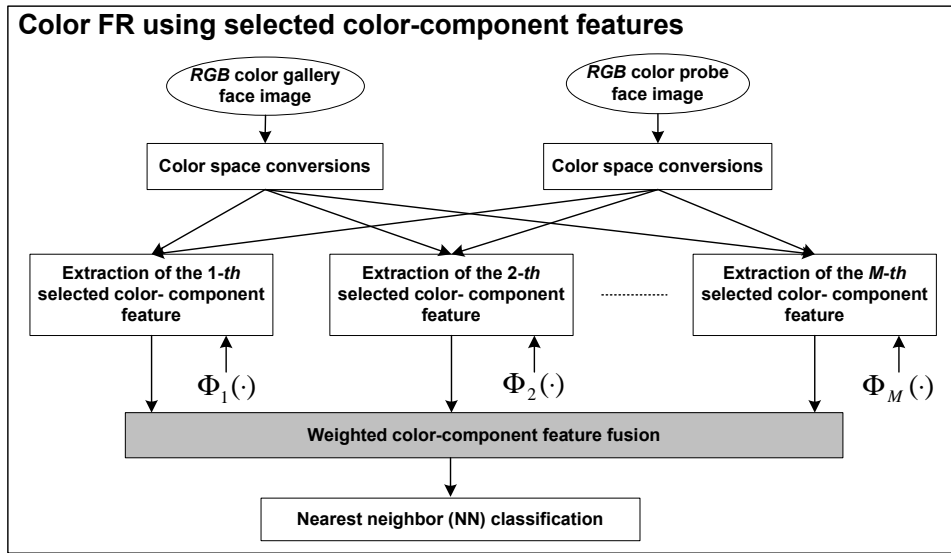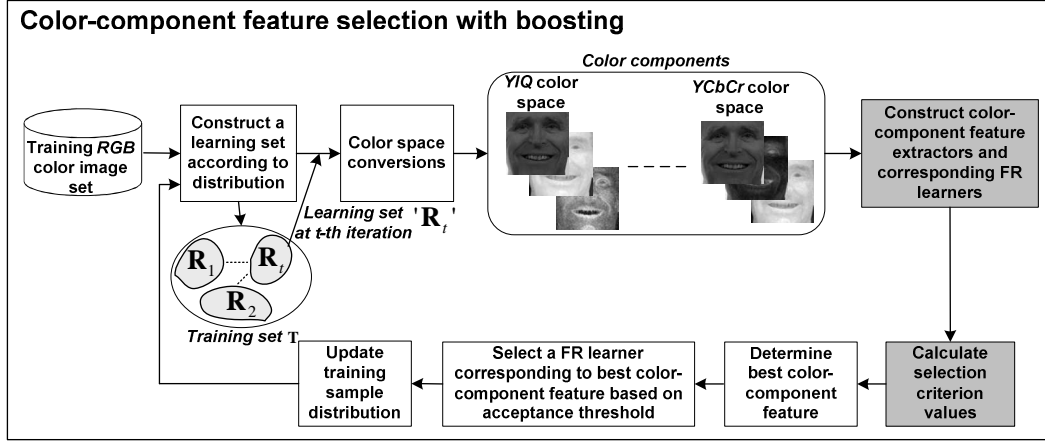
Accepted Manuscript for Publication in IEEE Transaction on Image Processing

(a)

(b)

Fig. 1. Proposed color FR framework. (a) Note that a learning set $\mathbf{R}_t$ is "most informative (or difficult)" subset of training samples given the currently selected color-component features. At each boosting round 't', the best color-component feature for classifying a learning set $\mathbf{R}_t$ is determined. (b) The $\{\Phi_t\}_{t=1}^{M}$ denote a set of M different feature extractors associated with M selected color-component features.

knowledge, our work is the first attempt to incorporate feature selection scheme underpinning boosting learning into FR methods using color information.

Fig.1 shows overall framework of the proposed color FR method which largely consists of two parts: 1) color-component feature selection with boosting and 2) color FR solution using selected color

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

component features. To determine the best color component feature at each boosting round for recognizing the hard-to-classify sample subset of a training set, termed "learning set", the effective selection criterion is proposed. The proposed selection criterion is in the form of penalty-based objective function with its associated weighting parameter for the purpose of selecting color-component features which not only produce small classification errors, but also keep their mutual dependence low. As demonstrated by the experimental results in Section 4.3, the proposed selection criterion is highly useful for achieving a low generalization classification error. In addition, to perform color FR, the color-component features chosen via our boosting framework are combined at the feature level. Specifically, selected color-component features are fused based on weighted feature fusion scheme−depending upon the associated confidence of each color-component feature−for achieving better FR performance.

In order to evaluate the effectiveness of the proposed color FR method, comparative and extensive experiments have been carried out. For this, five public face databases (DB) CMU-PIE [19], Color FERET [20], XM2VTSDB [21], SCface [22], and FRGC 2.0 [24] are used. Experimental results show that the results of the proposed method are impressively better than the results of other state-of-the-art color FR methods over different FR challenges including highly uncontrolled illumination, moderate pose variation, and small resolution face images.

The remaining of the paper is organized as follows: Section 2 describes our color-color component feature selection method within boosting learning framework. In particular, this section details the proposed selection criterion. In Section 3, we explain the proposed weighted feature fusion approach to combining selected color-component features for a FR purpose. In Section 4, we present extensive and comparative experimental results that demonstrate the effectiveness of the proposed color FR method. Conclusions and directions for future research are presented in Section 5.

2. Boosting color-component feature selection

In this paper, a multiclass boosting "Adaboost.M2" [12] framework is adapted to implement color-component feature selection. Differing from other boosting learning frameworks, the key advantage of

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

Adaboost.M2 framework is to force the weak learners to concentrate not only on the hard instances (or patterns), but also on the incorrect class labels that are hardest to classify [12]. As such, this boosting framework would fit well into our color-component feature selection that is devised for FR belonging to multi-classification problems. Further, this boosting framework is more flexible because the determination of error bound of the final hypothesis is free of the requirement that every weak hypothesis should have classification error less than 1/2 [12].

We now present the proposed color-component feature selection procedure. Let $\mathbf{L} = \{1,\ldots,C\}$ be the class label set, where C denotes maximum class label (or the number of classes). Also let $\mathbf{T}$ be a training set composed of N red-green-blue (*RGB*) color face images each denoted by $\mathbf{x}_t^{(i)}$ ($i = 1\ldots, N$) of size $H \times W$ pixels with a corresponding class label ℓ_i, where $\ell_i \in \mathbf{L}$. For each of the *RGB* color images in $\mathbf{T}$, color conversion can be done from the *RGB* color space to a number of different prespecified color spaces. Assuming that a total of '*K*' different color components are yielded from the color conversions under consideration, we then denote the *m*-th color component by f_m (e.g., C_b or C_r from YC_bC_r color space) comprising a color-component pool denoted by $\mathbf{F}$ for which $f_m \in \mathbf{F}$.

In our method, the best color-component feature (at each boosting round) for classifying a weighted version of $\mathbf{T}$ (i.e., weighted training samples) is determined based on selection criterion. To maintain a set of weights over the $\mathbf{T}$, the distribution denoted by $D_t(i)$ [12] for each training sample $\mathbf{x}_t^{(i)}$ can be determined at every boosting round. Initially, values of $D_t(i)$ are set equally, but on each round, they are newly updated in such a way that weak learners is forced to focus on the hard training samples.

Before presenting the entire color-component feature selection algorithm proposed, the way of constructing weak learners (so-called FR learners) and the proposed selection criterion will be described in detail in the following subsections.

2.1. Construction of FR learners

To construct weak learners at each round, the learning set $\mathbf{R}_t$ ($\mathbf{R}_t \subset \mathbf{T}$) is formed by choosing the

r hardest-to-classify training samples per class from the $\mathbf{T}$ according to distribution $D_t(i)$. Subsequently, a corresponding feature extractor Φ_m is constructed using $\mathbf{R}_t$ along with the m-th color-component f_m $(m=1,\ldots,K)$. Specifically, to form Φ_m, the Φ_m is trained with a set composed of the m-th color-component images that are generated from $\mathbf{R}_t$ via an associated color conversion. Here, $\Phi_m : \Re^{H \times W} \to \Re^J$ takes as input the m-th color component image of size $H \times W$ pixels and produces as output a corresponding J-dimensional feature. It is important to note that Φ_m can be obtained using any face feature extraction algorithm (e.g., using global- or local-based feature methods [14]). As opposed to conventional learners in the original boosting methods [12-13], the learning focus in our method lies on the *feature extractor* rather than the classifier. As such, it differs from the original boosting design in that only pure classifiers are generally used as the weak learners (termed "FR learner" hereafter) without considering feature extraction. This is mainly because the training process of typical FR algorithms is with emphasis on constructing a feature extractor instead of an associated classifier [14], [18].

Using Φ_m, a *corresponding FR learner* $h_{t,m}$ (at the t-th boosting round) is defined as follows:

$$h_{t,m}(\mathbf{x},n) = \frac{(d_{\max} - d_m^{(n)})}{(d_{\max} - d_{\min})},\tag{1}$$

where $d_m^{(n)} = S(\Phi_m(\mathbf{x}), \Phi_m(\mathbf{x}_g^{(n)}))$, $S(\cdot)$ is a metric function that measure distance between two input vectors in J-dimensional feature subspace, $\mathbf{x}$ and $\mathbf{x}_g^{(n)} (n \in \mathbf{L})$ denotes the *RGB* color image to be recognized and the n-th enrolled *RGB* color gallery image, respectively, $d_{\max} = \max(\{d_m^{(n)}\}_{n=1}^C)$, and $d_{\min} = \min(\{d_m^{(n)}\}_{n=1}^C)$. Note that in (1), outputs of FR learner $h_{t,m}(\cdot)$ are in the range $[0,1]$ for which it has the form $h_{t,m} : \Re^{H \times W} \times \mathbf{L} \to [0,1]$. Hence, it can satisfy the functional requirement of the general boosting algorithm [13] (including AdaBoost.M2 [12]), representing the degree (or confidence) for labeling $\mathbf{x}$ as the class n, namely, class label estimation.

2.2. Proposed selection criterion

At each boosting round, the best FR learner (i.e., the best color-component feature) should be determined from among K constructed FR learners $h_{t,m}$ $(m=1,\ldots,K)$, each of which depends upon a single color-component feature. To this end, a selection criterion plays a crucial role in determining the 'goodness' of feature selection. Referring to [11], in ensemble classification (including boosting), it has been shown that, to achieve the lowest *generalization error*, we need to create ensembles (or classifiers) with low training classification error, while at the same time their mutual dependence should be kept minimal. In particular, in our feature selection problem, mutual dependence between color-component features have to be carefully considered as different color channels may have similar properties from the view-point of classification. For instance, the *V* and *G* channels (from *HSV* and *RGB* color spaces, respectively) both encode the intensity information for green colors. Therefore, before a FR learner is selected, mutual dependence between the *new FR learner* and *each of the selected FR learners* should be examined to ensure that the complementary information (that improves classification) carried by the new FR learner is not captured by the preceding FR learners before.

To address the aforementioned issue, we develop an effective selection criterion which aims at making optimal balance between classification error and the degree of mutual dependence among selected FR learners. Here, using (1), classification error for $h_{t,m}$ is calculated based on 'pseudo-loss' [12]:

$$\varepsilon_{h_{t,m}} = \frac{1}{2}\sum_{i=1}^{N}D_t(i)\Big(1 - h_{t,m}(\mathbf{x}_t^{(i)},\ell_i) + \sum_{\ell\neq\ell_i}r_t(i,\ell)\,h_{t,m}(\mathbf{x}_t^{(i)},\ell)\Big) \tag{2}$$

where $r_t(i,\ell)$ is the mislabel weight vector (for details on the computation and implication of $r_t(i,\ell)$, please refer to [12]). Note that for computing $\varepsilon_{h_{t,m}}$, both *hard-to-classify* samples and *hard-to-separate* pairs of class labels are taken into account at the same time [12].

To quantify mutual dependence, one natural choice is to compute 'mutual information' in the theory of probability between pairs of the outputs of FR learners. However, this method requires accurate estimation of joint distributions over the outputs of FR learners; joint probabilities need to be computed

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

over $|\mathbf{L}|^2$ bins, where $|\mathbf{L}|$ denote the number of class labels. Hence, measuring mutual information may be ineffective in terms of simplicity for implementation, especially for general FR problems which often have to handle a large number of classes (usually hundreds or thousands of subjects). To cope with this, a simple but effective alternative way to measure mutual dependence between a pair of two FR learners has been devised as follows:

$$I(h_a, h_b) = \frac{1}{N} \sum_{i=1}^{N} \delta(h_a, h_b, \mathbf{x}_t^{(i)}), \tag{3}$$

where $\delta(\cdot)$ denotes an indicator function that returns 'one' only when, for i-th training sample, the class label predicted by h_a is equal to that predicted by h_b, otherwise returns zero and $0 \le I(h_a, h_b) \le 1$. Note that in (3), mutual dependence can be measured by investigating *agreement* of the outputs of two FR learners considered. Based on (3), we then examine the mutual dependence between the candidate FR learner $h_{t,m}$ (i.e., candidate color-component feature) and the already selected FR learners to avoid redundant FR learners in the following way:

$$\pi_{h_{t,m}} = \max_{h_{\hat{t}}} I(h_{t,m}, h_{\hat{t}}), \quad h_{\hat{t}} \in \mathbf{H}_M, \tag{4}$$

where $\mathbf{H}_M$ denotes the set of the best M selected FR learners before current boosting round.

Using (2) and (4), the best FR learner (at the t-th boosting round) is determined as follows:

$$h_t = \arg \min_{h_{t,m}} J_{h_{t,m}}, \tag{5}$$

where

$$J_{h_{t,m}} = \varepsilon_{h_{t,m}} + \lambda \pi_{h_{t,m}}, \tag{6}$$

and $0 \le \lambda < 1$. Note that in (6), objective function value $J_{h_{t,m}}$ of each $h_{t,m}$ has a penalty term $\pi_{h_{t,m}}$ and a weighting parameter λ that controls a trade-off between $\varepsilon_{h_{t,m}}$ and $\pi_{h_{t,m}}$ in order to enforce low mutual dependence between selected FR learners. Considering FR performance in our experiments, a good compromise has been found by setting λ in the range of [0.3, 0.5].

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

---

0. (Input)

(1) Color-component pool $\mathbf{F} = \{ f_m, m = 1,\dots,K \}$;

(2) A set of training *RGB* color images $\mathbf{T} = \{(\mathbf{x}_t^{(i)}, \ell_i)\}_{i=1}^{N}$ with class labels $\ell_i \in \mathbf{L}$, where $\mathbf{L} = \{1,\dots,C\}$;

(3) The acceptance threshold ζ^*;

(4) Total number of boosting rounds T;

1. (Initialization)

(1) Distribution $D_0(i) = 1/N$, for $i = 1,\dots,N$;

(2) Weight vector $w_1(i, \ell) = D_0(i)/(|\mathbf{L}| - 1)$, for $i = 1,\dots,N, \ell \in \mathbf{L} - \{\ell_i\}$;

(3) $M = 0$ and $\mathbf{H}_0 = \{\phi\}$;

2. (Repeat for $t = 1,\dots,T$)

(1) Set the mislabel weight vector for each training sample:

$$r_t(i, \ell) = \frac{w_t(i, \ell)}{\sum_{\ell \neq \ell_i} w_t(i, \ell)}, \text{ for } i = 1,\dots,N, \ell \in \mathbf{L} - \{\ell_i\};$$

(2) Update the distribution for each training sample:

$$D_t(i) = \frac{\sum_{\ell \neq \ell_i} w_t(i, \ell)}{\sum_{i=1}^{N} \sum_{\ell \neq \ell_i} w_t(i, \ell)}, \text{ for } i = 1,\dots,N, \ell \in \mathbf{L} - \{\ell_i\};$$

(3) Select 'r' hardest training samples per class according to $D_t(i)$ to form a learning set $\mathbf{R}_t$ ($\mathbf{R}_t \subset \mathbf{T}$);

(4) For $m = 1,\dots,K$

· Train a face feature extractor Φ_m using $\mathbf{R}_t$ along with the *m*-th color component (i.e., f_m);

· Construct a FR leaner $h_{t,m}$ based on Φ_m using Eq. (1) and get back a corresponding FR learner $h_{t,m} : R^{H \times W} \times L \rightarrow [0,1]$;

· Using Eq. (2), calculate the classification error $\varepsilon_{h_{t,m}}$ produced by $h_{t,m}$ over the whole weighted training set $\mathbf{T}$;

· Calculate 'mutual dependence degree' $\pi_{h_{t,m}}$ using Eq. (4) between $h_{t,m}$ and already selected FR learners included in a set $\mathbf{H}_M$;

(5) Determine the best FR learner h_t according to selection criterion with Eq. (5) and Eq. (6);

(6) Define $\varepsilon_t = \varepsilon_{h_t}$ and set $\beta_t = \varepsilon_t/(1-\varepsilon_t)$. If $\beta_t = 0$, then abort loop;

(7) Add the best FR learner at the *t*-th boosting round to the selected FR learner set if and only if $J_{h_t} < \zeta^*$;

· $M \leftarrow M + 1$;

· $\mathbf{H}_M = \mathbf{H}_{M-1} \cup \{h_t\}$;

(8) Update the weight vector:

$$w_{t+1}(i, \ell) = w_t(i, \ell) \beta_t^{\frac{1}{2}\left(1 + h_t(\mathbf{x}_t^{(i)}, \ell_i) - h_t(\mathbf{x}_t^{(i)}, \ell)\right)} \text{ for } i = 1,\dots,N, \ell \in \mathbf{L} - \{\ell_i\};$$

3. (Output)

M pairs of associated color component and feature extractor $\{(f_t, \Phi_t)\}_{t=1}^{M}$ of the chosen M FR learners $h_t \in \mathbf{H}_M$.

Fig. 2. Proposed algorithm for color-component feature selection.

2.3. Summary for color-component feature selection algorithm

At the *t*-th boosting round, the corresponding best FR learner h_t (determined through using (5) and (6)) is then added to $\mathbf{H}_M$ at a time, subject to condition that objective function value J_{h_t} for h_t (such that $J_{h_t} = \min_{h_{t,m}} J_{h_{t,m}}$) should be lower than predefined acceptance threshold ζ^*. It should be noted that the purpose of setting ζ^* is to provide better generalization classification error (see Section 4.3 for

further details).

The proposed color-component feature selection algorithm is summarized in Fig. 2. After terminating our boosting feature selection, M pairs of $\{(f_t, \Phi_t)\}_{t=1}^{M}$ associated with the chosen M FR learners contained in final $\mathbf{H}_M$ are used for performing color FR discussed in the next section.

3. Color FR using selected color-component features

Following from the original design of boosting framework [12], given M FR learners $h_t \in \mathbf{H}_M$ as its output, the traditional form of the final classifier for recognizing an unknown probe $\mathbf{x}_p$ is:

$$h_{\text{com}}(\mathbf{x}_p) = \arg\max_{n \in \mathbf{L}} \sum\nolimits_{h_t \in \mathbf{H}_M} \log(1/\beta_t) h_t(\mathbf{x}_p, n). \tag{7}$$

Note, by using (1), $h_t(\mathbf{x}_p, n)$ can be readily computed and the confidence parameter β_t is a function of ε_t as described in Step 2.(6) in Fig. 2. In (7), $h_{\text{com}}(\mathbf{x}_p)$ takes a weighted majority vote of the class predictions of M selected FR learners.

However, it should be emphasized that the ultimate goal of our boosting framework discussed in Section 2 is to select a subset of *color-component features* for achieving the best FR performance. Further, the learning focus of FR learners in this paper is stressed on *feature extractors* each constructed with a learning set of corresponding color-component images. Moreover, in the areas of multimodal fusion, fusing multiple features of the same biometric (or pattern) at the feature level can generally show better classification result than fusion methods working on other levels [15]. Based on the facts mentioned above, we decide to combine M selected color-component features at the feature level to achieve better FR performance. The following subsection provides a detailed description of proposed fusion method.

3.1. Weighted color-component feature fusion

Given M pairs of color component and associated feature extractor $\{(f_t, \Phi_t)\}_{t=1}^{M}$ of the chosen M FR learners $h_t \in \mathbf{H}_M$, the low-dimensional features of $\mathbf{x}_p$ and $\mathbf{x}_g^{(n)}$ along with f_t are obtained as follows (using the corresponding Φ_t):

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

$$\mathbf{f}_p^{(t)} = \Phi_t(\mathbf{x}_p) \quad \text{and} \quad \mathbf{f}_g^{(n,t)} = \Phi_t(\mathbf{x}_g^{(n)}), \tag{8}$$

where $\mathbf{f}_p^{(t)}, \mathbf{f}_g^{(n,t)} \in \Re^J$ and $t = 1, \ldots, M$. In order to generate the combined features for $\mathbf{x}_p$ and $\mathbf{x}_g^{(n)}$, M weighted low-dimensional features, given by (8), are combined at the level of the features (by stacking low-dimensional features in column order):

$$\mathbf{f}_p = \left[\log(1/\beta_1)(\mathbf{f}_p^{(1)})^{\mathrm{T}} \cdots \log(1/\beta_M)(\mathbf{f}_p^{(M)})^{\mathrm{T}} \right]^{\mathrm{T}}$$

$$\tag{9}$$

$$\text{and} \quad \mathbf{f}_g^{(n)} = \left[\log(1/\beta_1)(\mathbf{f}_g^{(n,1)})^{\mathrm{T}} \cdots \log(1/\beta_M)(\mathbf{f}_g^{(n,M)})^{\mathrm{T}} \right]^{\mathrm{T}},$$

where T denotes the transpose operator of a matrix and $\mathbf{f}_p, \mathbf{f}_g^{(n)} \in \Re^{MJ}$. Note that, for the same representation format, $\mathbf{f}_p^{(t)}$ and $\mathbf{f}_g^{(n,t)}$ should be individually normalized to have zero mean and unit variance before concatenation [15]. In (9), depending on the confidence (in terms of classification accuracy) of M color-component features chosen through boosting, we give some color-component features more weight than others when computing distance between $\mathbf{f}_p$ and $\mathbf{f}_g^{(n)}$. As such, weighted feature fusion shown in (9) allows for effectively facilitating a *complementary effect* between its different components, leading to positively affect the classification performance.

To perform FR tasks (identification or verification) on $\mathbf{x}_p$, a NN classifier is then applied to determine the identity of $\mathbf{x}_p$ as follows:

$$l(\mathbf{x}_p) = l(\mathbf{x}_g^{(n^*)}) \quad \text{and} \quad n^* = \arg\min_{n \in \mathbf{L}} S(\mathbf{f}_p, \mathbf{f}_g^{(n)}), \tag{10}$$

where $l(\cdot)$ returns the identity label of a face image.

4. Experiments

Five public CMU-PIE [19], Color FERET [20], XM2VTSDB [21], SCface [22], and FRGC 2.0 [24] face DBs were used to evaluate the proposed method. All facial images used in our experiments were

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

(a)

(b)

$-45°$ $-22.5°$ $0°$ $22.5°$ $45°$

(c)

(d)

(e)

Fig. 3. (a) Examples of the facial images with flash illumination from the CMU-PIE DB. (b) Examples of facial images with uncontrolled illumination condition from the XM2VTSDB. (c) Examples of facial images with pose variations from the Color FERET DB. (d) Examples of facial images with variation in face resolutions from the SCface DB. (e) Example of facial images from the FRGC 2.0 DB.

manually cropped from original images based on the locations of the two eyes. Each cropped facial image

was rescaled to the size of 64 x 64 pixels (see Fig. 3).

To construct a face feature extractor $\Phi_m(\cdot)$ (described in Section 2), four popular low-dimensional

feature extraction techniques were used: Principal Component Analysis (PCA) [16], Fisher's Linear

Discriminant Analysis (FLDA) [17], Regularized Linear Discriminant Analysis (RLDA) [18], and

Enhanced Fisher linear discriminant Model (EFM) [3]. As for the $S(\cdot)$ in (1) and (10), the Euclidean

distance was used for FLDA and RLDA, while the Mahalanobis and cosine distance measures for PCA

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

and EFM, respectively. In addition, in all experiments, the frontal-view images with neutral illumination and expression were used to build the gallery set.

To form the color-component pool $\mathbf{F}$ used for color-component feature selection process described in Fig. 2, the 36 different color components from the following 12 different color spaces were used[1]: 'YC_bC_r', 'YIQ', 'HSV', 'JPEG-XR' (C_o, C_g, and Y), 'RGB', 'YUV', 'XYZ', 'CIE $L^*a^*b^*$', 'YP_bP_r', 'YD_bD_r', 'CIE $L^*u^*v^*$', and 'CIE L^*ch'. Note that the color spaces used can be derived from the RGB color space by means of either linear or nonlinear transformations. A detailed description for the color spaces used can be found in [8].

4.1. Evaluation of our method under different challenges

In this section, we present comparative experimental results to demonstrate the effectiveness of our method under different FR challenges. For comparison purpose, the following state-of-the art color FR methods are implemented: "Hybrid Color and Frequency Feature" (CFF) method [4], "Color Space Normalization" (CSN) method [7], "Color Image Discriminant Model" (CID) method [2], "Independent Color Space" (ICS) method [3], "Hybrid Color Configuration RQC_r" method [5].

For the CFF method, the hybrid 'RIQ' color space was used as proposed in [4]. In addition, as recommended by [4], the same size of masks used to select frequency sets in frequency domain was used. For the CSN method, the normalized hybrid 'ZRG' color space using across-color-component normalization technique [7] was used as this method achieves the best FR performance of all normalized color spaces evaluated in [7]. For the CID method, we implemented its extended version based on the RGB color space [2]. In addition, following the same parameter values as used in [2], the initial value of the CID algorithm and convergence threshold were set to '[1/3,1/3,1/3]' and '0.1', respectively. The ICS defines statistically independent component images that are created using a blind source separation technique; in our experiment, Comon's ICA algorithm [23] was used to compute mutual information

---

[1] The color space conversion is performed using the "Color Space Converter" Matlab toolbox available at: http://www.mathworks.com/matlabcentral/fileexchange/7744.

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

Table 1. Rank-one identification rates (%) [20] obtained for six different color FR methods on the 'CMU-PIE' and the 'XM2VTSDB' face images having illumination variation. Note that, for the proposed method, four hardest training images per subject (i.e., $r = 4$) are selected from a training set at each boosting round to form a learning set.

Color FR method	Used color information fusion method	Feature extraction method		
		PCA	FLDA	RLDA
Proposed	FL fusion with different weights	**82.27 ±1.01**	**93.85 ±2.45**	**98.27 ±1.17**
Proposed	FL fusion with uniform weights	78.34 ±0.11	88.05 ±0.92	93.30 ±1.54
CFF	Weighted similarity score fusion [4]	72.41 ±1.98	84.97 ±0.83	88.57 ±1.05
CSN	IL fusion [7]	74.73 ±1.11	88.97 ±2.42	92.57 ±1.43
ICS	IL fusion [3]	68.98 ±2.99	79.23 ±2.03	83.56 ±1.77
CID	IL fusion [2]	73.58 ±2.56	86.23 ±3.42	89.36 ±2.11
RQCr	IL fusion [5]	70.87 ±2.19	81.32 ±0.32	85.00 ±1.01

Note that FL correspond to the feature-level, while IL to the input-level. In addition, bold value denotes the best result of FR approaches in each low-dimensional feature extraction technique and the similar notations are also used in the following tables.

Table 2. Rank-one identification rates (%) obtained for six different color FR methods on the 'Color FERET' face images having pose variation. Note that, for the proposed method, we set $r = 3$ to form a learning set.

Color FR method	Used color information fusion method	Feature extraction method		
		PCA	FLDA	RLDA
Proposed	FL fusion with different weights	**69.90 ±1.61**	**84.09 ±2.63**	**87.53 ±2.78**
Proposed	FL fusion with uniform weights	67.31 ±1.68	80.92 ±1.77	83.24 ±1.86
CFF	Weighted similarity score fusion [4]	54.30 ±0.65	69.90 ±1.37	72.80 ±0.75
CSN	IL fusion [7]	64.53 ±0.59	77.10 ±1.50	80.14 ±0.46
ICS	IL fusion [3]	51.56 ±1.68	59.58 ±0.75	69.52 ±1.79
CID	IL fusion [2]	52.41 ±3.01	71.45 ±2.68	70.38 ±3.11
RQCr	IL fusion [5]	49.79 ±1.86	61.12 ±2.44	66.76 ±0.89

Table 3. Rank-one identification rates (%) obtained for six different color FR methods on the 'SCface' images having spatial resolution variation. Note that, for the proposed method, we set $r = 3$ to form a learning set.

Color FR method	Used color information fusion method	Feature extraction method		
		PCA	FLDA	RLDA
Proposed	FL fusion with different weights	**42.56**	**56.06**	**62.78**
Proposed	FL fusion with uniform weights	38.81	51.68	59.09
CFF	Weighted similarity score fusion [4]	28.88	39.31	43.55
CSN	IL fusion [7]	34.86	48.70	53.51
ICS	IL fusion [3]	26.78	37.78	41.89
CID	IL fusion [2]	37.76	49.54	54.70
RQCr	IL fusion [5]	34.85	46.34	49.61

between color components and high-order statistics as suggested in [3]. The ICS shows the best FR performance of all color image representations evaluated in [3]. In [5], the authors show that the hybrid *RQCr* color configuration is considerably effective for recognizing low-resolution face images. In

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

addition, in the proposed method, the acceptance threshold ζ^* in Fig. 2 and the weighting parameter λ in (6) were empirically set to 0.6 and 0.4, respectively, across all experiments in this section. However, it should be emphasized that varying ζ^* in the range of [0.55, 0.65] did not much alter the FR performance of the proposed method in our experiments. Further, in order to validate the advantage of using different weights for all selected color components as proposed in (9), we also present the results using uniform weights (i.e., all components have the same weights when computing distance).

Comparative experimental results under variations of illumination, pose, and spatial resolution are presented as follows:

1) **Under illumination variation**: We compare the robustness of the proposed color FR with other color FR methods against extensive variations in illumination using CMU-PIE and XM2VTSDB face DB. In this experiment, 1,428 frontal images of 68 subjects (21 images per subject) were collected from the CMU- PIE; the facial images for each subject have 21 different illumination variations (using the 'room lighting off' condition). From the XM2VTSDB, 900 frontal images of 100 subjects were obtained; each subject included nine facial images captured with no control on severe illumination variations. Fig. 3(a) and Fig. 3(b) show examples of facial images used in this experiment. By using random partition, the training set consisted of (6 images x 168 subjects), while the remaining 1,320 images were used to create a probe set. Table 1 shows the rank-one identification rates of six different color FR methods. To guarantee stable experimental results, 20 independent runs of aforementioned random partitions were executed. Thus, all results presented in Table 1 were averaged over 20 runs. The results show that our method (using different weights) outperforms the other five color FR methods for all feature extraction techniques. For instance, for the case of using RLDA, our method can attain about 14.71%, 12.71%, 9.70%, 8.91% and 7.70% improvement, compared to 'ICS', 'RQC_r', 'CFF', 'CID' 'CSN' methods, respectively.

2) **Under pose variation**: We further assess the usefulness of the proposed method under moderate

pose variations. A total of 1,378 face images of 107 subjects were collected from the Color FERET face DB. It should be noted that the rotated face images that both eyes can be reliably identified for normalization were only collected. The facial images used include five different pose angles ranging from $-45°$ to $+45°$ (see Fig. 3(c)). Also note that all the images have neutral expression and illumination. By using random partition, the training set consisted of (5 images x 107 subjects), while the probe set contained the remaining 843 images of the same 107 subjects. The comparison results are described in Table 2. It is shown that the proposed method attains the highest recognition accuracies for all feature extraction methods followed by the 'CSN', the 'CFF', and the 'CID' color FR methods. In particular, compared to the second best method 'CSN', identification rates can be improved by 5.37%, 6.99%, and 7.39% for PCA, FLDA, and RLDA, respectively. This demonstrates the effectiveness of our method under pose variations.

3) **Under spatial resolution variation**: In this experiment, we further evaluate the effectiveness of the proposed method against small resolution face images. To this end, 2,080 face images of 130 subjects were selected from SCface DB. This dataset has been designed to test the FR algorithms in *real-world surveillance* setup [22]. Different quality face images were taken with 5 different commercially available surveillance cameras of various quality and resolution. There are three images per subject for each camera, captured at three different distances (4.20, 2.60, and 1.00 m) as described in [22]. As shown in Fig. 3(d), some of these surveillance images are of extremely low quality and resolution. In real-life surveillance-like FR applications, it is reasonable to assume that high-resolution face images are chosen as training and gallery images. On the other hand, the probe to be tested may have lower and various face resolutions [5]. Hence, in our experiments, gallery set consisted of one frontal mug shot (per subject) image, and the training set consisted of 5 face images with distance label '3' per subject (i.e., captured with 5 different surveillance cameras at a distance of 1.00 m), while the probe set consisted of 10 face images (per subject) with distance labels '1' or '2' (i.e., captured at a distance 2.60 and 4.20 m). Note that, to match a low-resolution

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

probe to a high-resolution gallery face, the probe has been upsampled to be matched with template

size of 64 x 64 pixels by using a cubic interpolation technique before FR. The experimental results

of six different color FR methods on the varying face resolutions are described in Table 3. From

Table 3, we can see that recognizing face images (with much lower resolution) collected from

'SCface' DB is significantly challenging. The rank-one identification rate averaged over all of the

color FR and feature extraction methods is less than around 45%. However, it should be noted that

the proposed method can achieve the best FR performance of up to around 63% when using RLDA.

4.2. Comparison with other color FR methods

In this section, we conduct comparative experiments on the FRGC 2.0 dataset to further evaluate our

method. Here, the "FRGC Experiment 4" is chosen to assess the proposed method because the FRGC

Experiment 4 has been reported to be the most challenging FRGC experiment [3], [24]. Note that direct

comparisons are made with other state-of-the-art results reported recently by other researchers on these

FRGC 2.0 dataset. As such, all the results for comparison are directly cited from papers published recently.

Note that, in this experiment, the performance measurement is Face Verification Rate (FVR) at False

Accept Rate (FAR) equal to 0.1%, which corresponding to ROC III curve [24]. In addition, note that the

EFM is adopted for a face feature extractor in our method because all other methods (to be compared)

make use of EFM for extracting low-dimensional features. From the comparison shown in Table 4, we

can see that our method has made impressive improvements, which further validate the effectiveness of

our method.

Table 4. Comparisons with other state-of-the-art color FR methods on the "FRGC 2.0 Experiment 4". Note that z-score normalization is used to compute FVR and FAR. Also note that, for our method, we set $r = 7$ to form a learning set.

Color FR Method	FVR (ROC III) when FAR = 0.1%
Proposed (using different weights)	**86.97%**
CFF in [4]	80.30%
CSN in [7]	72.86%
ICS in [3]	73.69%
CID in [2]	78.26%

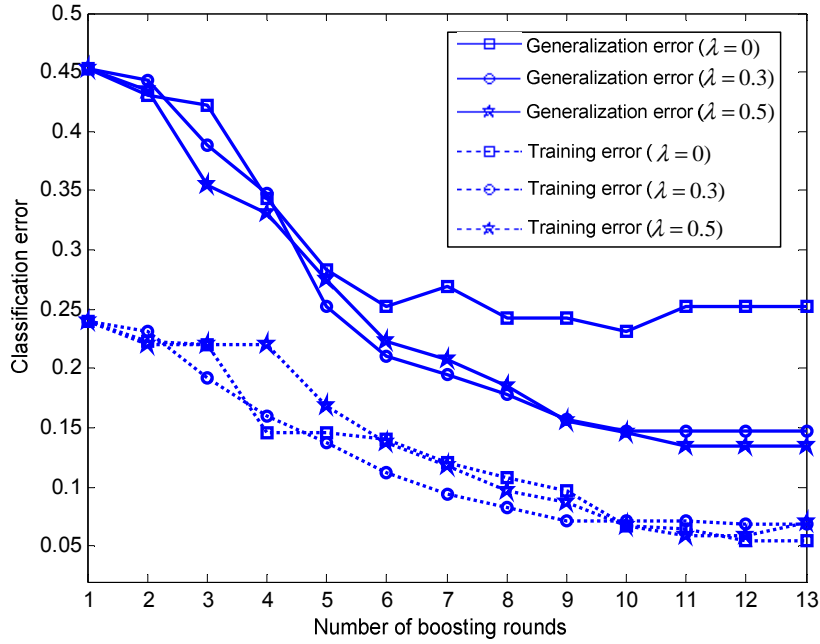Accepted Manuscript for Publication in IEEE Transaction on Image Processing

Fig. 4. Impact of weighting parameter λ on generalization classification performance. Note that RLDA was used for feature extractor and the acceptance threshold values for all of the three different weighting values were set to 0.6 for fair comparisons.
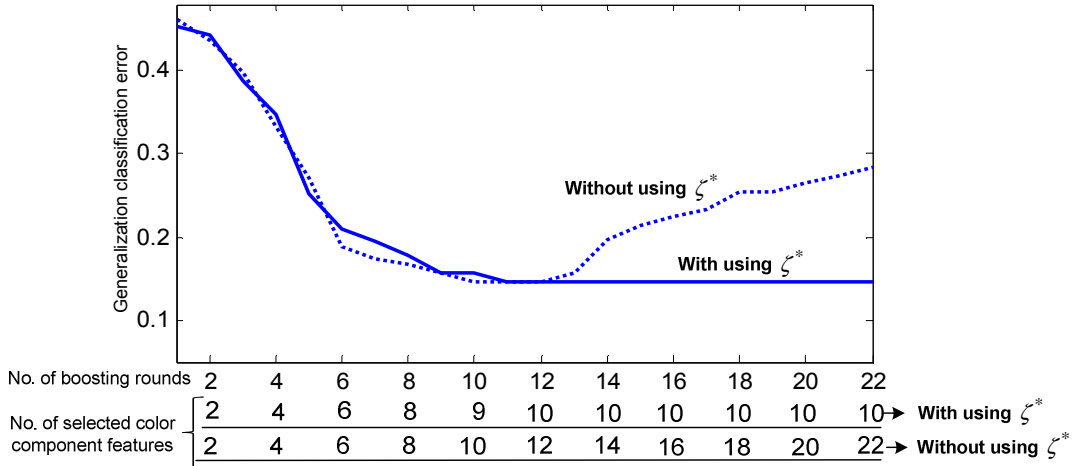

Fig. 5. Effectiveness of using acceptance threshold ζ^* on selecting an optimal subset of color-component features. Note that RLDA was used for feature extractor and, for the case of using ζ^*, we set $\zeta^* = 0.6$.

4.3. Effectiveness of weighting parameter and acceptance threshold

Note that in (6), the weighting parameter λ has been introduced aiming to consider both mutual dependence between selected color-component features and their classification errors in our boosting

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

based color-component feature selection. In addition, as described in Fig. 2, the acceptance threshold ζ^* is used to prevent the color-component features having either much lower classification errors or much higher mutual dependences from being selected. It should be noted that main objective of using both λ and ζ^* is to achieve better generalization classification (or recognition) performance. To validate the effectiveness of using λ and ζ^* in the proposed method, experimental analysis has been performed using the face image dataset collected from the Color FERET. A detailed description of the experimental dataset used is given in Section 4.1.

Fig. 4 shows the impact of weighting parameter λ on generalization classification performance of the proposed method. Note that in Fig. 4, the selected color-component features determined at each boosting round were used to compute the classification errors by using (9) and (10). Also note that $\lambda = 0$ means that the classification error defined in (2) is only taken into account during color-component feature selection process. As shown in Fig. 4, the training errors for all of the three different weighting parameter values can be reduced as the number of boosting rounds is increased and finally converged to nearly same constant value. However, for the case of generalization classification error, we can see that generalization errors for both $\lambda = 0.3$ and $\lambda = 0.5$ converge to much lower values (after passing the 10-*th* boosting round) than that obtained for $\lambda = 0$. This result ensures that considering both mutual dependence and classification errors by using λ for color-component feature selection allows achieving a low generalization error.

Fig. 5 shows the effectiveness of using acceptance threshold ζ^* on selecting an optimal subset of color-component features. In Fig. 5, the number of selected color-component features corresponding to each boosting round is also presented. As shown in Fig. 5, when using ζ^*, our method stops adding color-component features at boosting round (i.e., the 11-*th* boosting round) where the lowest generalization classification error has been attained. This is done by protecting addition of color-component features whose objective function values defined in (6) are lower than predetermined value of

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

ζ^*. On the other hand, when not using ζ^*, as more color-component features are added, generalized classification error decreases, which then arrives at minimum error (at the number of features 10 through 12), and eventually increases again. These results demonstrate that making use of ζ^* with appropriate value is useful for determining the best number of color-component features as well as the types of color-component features ("best" in the sense that generalization classification error is minimized).

5. Conclusions and future research

In this paper, a novel and effective color FR method is proposed. It is based on the selection of the best color-component features (from various color models) using the proposed variant of boosting learning framework. These selected color-component features are then combined into a single concatenated color feature using weighted feature fusion. Our results clearly demonstrate the effectiveness of the proposed method in terms of both absolute performance and comparative performance against state-of-the-art color FR methods.

In this paper, the extraction of color-component features is restricted to using global-based feature extraction methods (such as PCA and LDA). However, other face features (or descriptors) can be readily incorporated into the proposed selection framework, aiming to find the most suitable features for a given FR task. In particular, for the future work, we will extend our work by applying popular local-based feature extraction techniques, such as Gabor wavelets [14] or Local Binary Pattern (LBP) [25], to construct the FR learners defined in (1). For instance, color LBP (CLBP) feature extraction technique proposed in [26] can be easily applied to the construction of FR learners. This extension will allow for finding better color-component features via the proposed boosting feature selection algorithm in terms of achieving the best face recognition results. In addition, even though standard color spaces (such as *RGB* and *YC_bC_r*) are only considered during boosting feature selection process in this paper, our method will be readily extended by incorporating new color spaces [2-3], [7] (e.g., normalized '*ZRG*' color space proposed in [7]) devised for a color FR purpose. This is expected to yield better performance. In addition, for the future work, we will exploit which combinations of color components generalize well across the

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

illumination variations in the context of FR.

Acknowledgement

References

[1] P. Shih and C. Liu, "Improving the Face Recognition Grand Challenge Baseline Performance Using Color Configurations Across Color Spaces," *Proc. IEEE Int'l Conf. on Image Processing (ICIP)*, 2006

[2] J. Wang and C. Liu, "Color Image Discriminant Models and Algorithms for Face Recognition," *IEEE Trans. Neural Networks*, vol. 19, no. 12, pp. 2088-2097, 2008.

[3] C. Liu, "Learning the Uncorrelated, Independent, and Discriminating Color spaces for Face Recognition," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 2, pp. 213-222, 2008.

[4] Z. Liu and C. Liu, "A Hybrid Color and Frequency Features Method for Face Recognition," *IEEE Trans. Image Processing*, vol. 17, no. 10, pp. 1975-1980, 2008.

[5] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color Face Recognition for Degraded Face Images," *IEEE Trans. Systems. Man and Cybernetics-Part B*, vol. 39, no. 5, pp. 1217-1230, 2009.

[6] M. Rajapakse, J. Tan, and J. Rajapakse, "Color channel encoding with NMF for face recognition," *Proc. IEEE Int'l Conf. on Image Processing,* 2004.

[7] J. Yang, C. Liu, and L. Zhang, "Color space normalization: Enhancing the discriminating power of color spaces for face recognition" *Pattern Recognit.,* vol. 35, no. 1, pp. 615-625, 2010.

[8] R. Likac and K. N. Plataniotis, *Color Image Processing: Methods and Application.* New York: CRC, 2007.

[9] H. Stokman and T. Gevers, "Selection and Fusion of Color Models for Image Feature Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 371-381, 2007.

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

[10] R.L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 696-796, 2002.

[11] A. Mura, "Upper bounds for error rates of linear combinations of classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 591-602, 2002.

[12] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119-139, 1997.

[13] R. E. Schapire, "The boosting approach to machine learning: An overview," *MSRI Workshop Nonlinear Estimation and Classification*, pp. 149-172, 2002.

[14] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Trans. Image Processing*, vol. 18, no. 8, pp. 1885-1886, 2009.

[15] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.,* vol. 38, no. 12, pp. 2270-2285, 2005.

[16] M. A. Turk and A. P. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71-86, 1991.

[17] P. N. Belhumeur, J. P. Hesphanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern. Anal. Machine Intell.*, vo1. 9, no. 7, pp. 711-720, 1997.

[18] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognit. Lett.,* vol. 24, no. 16, pp. 3079-3087, 2003.

[19] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression Database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615-1618, 2003.

[20] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, no. 10, pp. 1090-1104, 2000.

[21] K. Messer, J. Mastas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," *Proc. IEEE Int'l Conf. on AVBPA*, 1999.

[22] M. Grgic, K. Delac, and S. Grgic, "SCface - surveillance cameras face database," *Multimedia Tools and Applications Journal*, 2009, DOI:10.1007/s11042-009-0417-2.

[23] P. Comon, "Independent component analysis, a new concept?," *Signal Process.,* vol. 36, pp. 287-314, 1994.

[24] P.J. Phillips, P.J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenges," *Proc. IEEE Int'l Conf. on*

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

Computer Vision and Pattern Recognition, 2005.

[25] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Pattern: Application to Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037-2041, 2006.

[26] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Using Color Local Binary Pattern Features for Face Recognition," *Proc. IEEE Int'l Conf. on Image Processing (ICIP)*, 2010.

Jae Young Choi received the B.S. degree from Kwangwoon University, Seoul, South Korea, in 2004 and the M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008. He is currently working toward the Ph.D. degree at KAIST. In 2007, he worked as an intern researcher for the Electronics and Telecommunications Research Institute (ETRI) in South Korea. In 2008, he was a visiting student researcher at the University of Toronto in Canada. His research interests include face recognition/detection/tracking, image/video indexing, medical image processing, pattern recognition, machine learning, the Social Web, and computer vision.

Yong Man Ro (M'92-SM'98) received the B.S. degree from Yonsei University, Seoul, South Korea, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. In 1987, he was a visiting researcher at Columbia University, and from 1992 to 1995, he was a visiting researcher at the University of California, Irvine and KAIST. He was a research fellow at the University of California, Berkeley and a visiting professor at the University of Toronto in 1996 and 2007, respectively. He is currently holding the position of full professor at KAIST, where he is directing the Image and Video Systems Lab. He participated in the MPEG-7 and MPEG-21 international standardization efforts, contributing to the definition of the MPEG-7 texture descriptor, the MPEG-21 DIA visual impairment descriptors, and modality conversion. His research interests include image/video processing, multimedia adaptation, visual data mining, image/video indexing, and multimedia security. Dr. Ro received the Young Investigator Finalist Award of ISMRM in 1992 and the Scientist Award in Korea in 2003. He served as a TPC member of international conferences such as IWDW, WIAMIS, AIRS, and CCNC, and he was the co-program chair of IWDW 2004."

Konstantinos N. (Kostas) Plataniotis received his B. Eng. degree in Computer Engineering from University of Patras, Greece in 1988 and his M.S and Ph.D degrees in Electrical Engineering from Florida Institute of Technology (Florida Tech) in Melbourne, Florida, in 1992 and 1994 respectively. He is a Professor with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto in Toronto, Ontario, Canada, an Adjunct Professor with the School of Computer Science at Ryerson University, Canada. He is the Director of the Knowledge Media Design Institute (KMDI) at the University of Toronto, and the Director of Research for U of T's

Accepted Manuscript for Publication in IEEE Transaction on Image Processing

Identity, Privacy and Security Institute. His research interests include biometrics, communications systems, multimedia systems, and signal & image processing. Dr. Plataniotis is the Editor in Chief (2009-2011) for the IEEE Signal Processing Letters, a registered professional engineer in the province of Ontario, and a member of the Technical Chamber of Greece. He is the 2005 recipient of IEEE Canada's Outstanding Engineering Educator Award "for contributions to engineering education and inspirational guidance of graduate students" and the co-recipient of the 2006 IEEE Trans. on Neural Networks Outstanding Paper Award for the published in 2003 paper entitled "Face Rscript for publication in IEEE Transactions on Image Processing