Neurocomputing 168 (2015) 566–574

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Adaptive weighted fusion: A novel fusion approach for image classification

Yong Xu [a,b,*], Yuwu Lu [a]

[a] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China
[b] Key Laboratory of Network Oriented Intelligent Computation, China

ARTICLE INFO

*Article history:*
Received 1 September 2014
Received in revised form
10 March 2015
Accepted 20 May 2015
Communicated by: Su-Jing Wang
Available online 27 May 2015

*Keywords:*
Adaptive weighted fusion
Image classification
Sum fusion

ABSTRACT

Score fusion is a very competent fusion approach and weighted score fusion is the most preferable score fusion approach. To automatically set proper weights is the most important key of weighted score fusion and it seems that there are no truly adaptive weighted fusion approaches at present. In this paper we design a perfect adaptive weighted fusion approach, which automatically determines optimal weights and no any manual setting is needed. Though the proposed approach is very simple and quite easy to implement, it can obtain better performance than previous state-of-the-art approaches.


## 1. Introduction

As fusion enables more information from multiple sources to be exploited and allows higher accuracy to be achieved, fusion has been used in various tasks such as video retrieval [1], cryptosystems [2], multi-biometrics [3–5] and spoken language recognition [6].

Fusion is usually performed at three levels, i.e. feature level, score level and decision level [7,8]. Fusions at these levels have different advantages. The decision level fusion is very easy to implement. However, the implementation of decision level fusion does not allow the multisource information to be fully exploited, because the decision contains only too little information from the original data. For example, in the biometrics-based personal verification, the decision to be fused is just a binary variable, with the value of "accept" or "reject". Fusion at the feature level integrates the multisource information at the most early stage, so it is able to exploit the most information of the original data [8]. However, fusion at the feature level should resolve the problem that different kinds of data might be inconsistent and incompatible. As a result, fusion at the score level is a good way [5,9,10].

Conventional score fusion approaches can be grouped into three kinds, i.e. transformation based score fusion, classifier based score fusion and density based score fusion [4]. In classifier based score fusion approaches, scores from different data sources are combined as a feature vector and a classifier is constructed to perform classification [11,12]. Besides conventional machine learning approaches such as multi-layer perceptron (MLP) have been exploited for classifier based score fusion [13], the boosting approaches such as the ones proposed in [14,15] have also shown good performance in this kind of fusion. The key issues of classifier based score fusion approaches are as follows. First, they are faced with the problem of the unbalanced training set. For example, in the personal verification, genuine match scores available for training are much fewer than impostor scores. Second, the cost of misclassification and classifier should be carefully selected. In transformation based score fusion approaches, the scores should be first transformed (normalized) to a common domain before they are integrated. Because the transformed (normalized) scheme is sample dependent, empirical evaluation is usually involved in the implementation of this kind of approaches [9,16]. Transformation based score fusion usually uses the sum rule, maximum rule, minimum rule and product rule to integrate the scores of different data sources [7]. Most of previous literatures show that the sum rule achieves promising performance [17]. If the score densities are evaluated accurately, density-based approaches seem to be able to achieve optimal performance. However, they suffer from the problem that they are so complex and it is hard to implement them. In particular, to model the density distributions has a very high complexity [18]. The Gaussian mixture model (GMM) based density estimation has been widely used in score fusion owing to its good theoretical properties [19]. However, to determine a suitable number of components for GMM is a challenging task [4].

Besides the above three kinds of score fusion approaches, there are also a few available ways to score fusion. For example, receiver operating characteristic (ROC) based score fusion approaches were

* Corresponding author at: Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China.
*E-mail address:* yongxu@ymail.com (Y. Xu).

http://dx.doi.org/10.1016/j.neucom.2015.05.070
0925-2312/

also proposed. Typical examples of this kind of score fusion approaches include the least square error based framework developed in [20], the margin based ranking [21], and the optimizing approach of area under the Curve (AUC) [22].

It is sure that weighted score level fusion can better take advantages of data from different sources [23–27]. However, it is hard to determine optimal weights for weighted score level fusion. It seems that there are still no automatic weight selection procedures and a number of previous weighted score level fusion approaches depend on only empirical selection of the weights. As a consequence, to design an automatic and adaptive weighted score level fusion approach is very crucial.

Some attempts have been made for improving weighted score level fusion. For example, Jain et al. proposed user-specific parameters for multi-biometrics [28]. Moumene et al. studied the estimation of fusion weights in the exposure fusion problem [29]. However, the algorithm depends on the Lagrange multipliers and it is not suitable for pattern classification tasks. The work presented in [28] can be viewed as user-specific adaptive fusion approach. Besides this kind of approaches, other adaptive fusion studies are also made [30–32]. For instance, the approach proposed in [32] is a typical test-sample-specific adaptive weighted fusion approach. In other words, in this approach the weights of different data sources vary with test samples. The sample quality based weighted score fusion approach proposed in [18] can be also viewed as a sample-specific adaptive weighted fusion approach. It has the idea that since the sample quality varies the weights should be various for different samples. It should be pointed out that most of the above these attempts cannot completely implement automatic determination of adaptive weights. Some of them estimate the weights by using "exhaustive" search and a certain error-rate based criterion [28].

This paper devises a perfect adaptive weighted fusion approach. The approach can automatically set proper weights for score fusion and no any manual setting is needed. Because the automatically selected weights are very reasonable, this approach can well integrate the advantages of complementary data sources. Extensive experiments demonstrate that the devised method outperforms previous state-of-the-art approaches.

The remainder of this paper is organized as follows. Section 2 reviews related works. Section 3 presents the proposed adaptive weighted fusion approach and its rationale and advantages. Section 4 shows the results of conducted extensive experiments. Section 5 provides the conclusions of the paper.

## 2. Related works

In this section, we make a brief review of typical score level fusion schemes. The score level fusion can use more information than the decision level fusion. Moreover, as the score level fusion allows multiple scores to be independently treated and integrated, it usually obtains a very high accuracy [7].

Kittler et al. proposed a theoretical framework of score level fusion for consolidating the evidence obtained from multiple classifiers. Under this framework, the score level fusion can be implemented by the sum rule, product rule, max rule, min rule, median rule or majority voting [17]. In order to use these rules, we should convert matching scores into posteriori probabilities. Kittler et al. considered the problem of classifying an input pattern $Z$ into one of $m$ possible classes based on the evidence provided by $R$ different classifiers. Let $x'_i$ be the feature vector which is derived from input pattern $Z$ presented to the *ith* classifier. $p(\omega_j|x'_i)$ is the posterior probability of pattern $Z$ belonging to class $\omega_j$ given feature vector $x'_i$. Let $c \in (1, 2, \cdots, L)$ be the class to which input pattern $Z$ is finally assigned. The following rules can be used to determine $c$:

(1) Product Rule: Assume representations $x'_1, x'_2, \cdots, x'_R$ are statistically independent. The input pattern is assigned to class $c$ such that

$$c = \arg\max_j \prod_{i=1}^{R} p(\omega_j|x'_i).$$

(2) Sum Rule: Besides the assumption of statistical independence of the multiple representations used in the product rule, the sum rule also assumes that posteriori probabilities computed by the individual classifiers do not deviate much from the prior probabilities. The sum rule assigns the input pattern to class $c$ such that

$$c = \arg\max_j \sum_{i=1}^{R} p(\omega_j|\boldsymbol{x'_i}).$$

(3) The max rule approximates the mean of the posteriori probabilities by the maximum value. The max rule assigns the input pattern to class $c$ such that

$$c = \arg\max_j \max_i p(\omega_j|x'_i).$$

(4) The min rule is derived by bounding the product of posteriori probabilities. The min rule assigns the input pattern to class $c$ such that

$$c = \arg\max_j \min_i p(\omega_j|x'_i).$$

Prabhakar and Jain argued that the assumption of statistical independence of the feature sets may not be true in a multimodal biometric system [33]. They showed that their method would be optimal in the Neyman–Pearson decision sense, if sufficient training data were available to estimate the joint densities. Bigun et al. proposed a new algorithm for the fusion module of a multimodal biometric system that took into account the estimated accuracy of the individual classifiers in the fusion procedure [34]. In [34], the authors used bayesian statistics when combining the scores of different biometric matchers. The authors showed that their multimodal system using image and speech data achieved higher recognition rates than individual modalities. In [35], the authors proposed a weighted sum fusion scheme to combine matching scores of individual matchers. In [36], the authors proposed the max-score and min-score approaches. In the max-score approach, the maximum value among the scores of individual matchers is regarded as the fusion result. In the min-score approach, the minimum value among the scores of individual matcher is the fusion result. Taigwan et al. [37] presented a deep face system which used for unconstrained face recognition. In [38], the authors proposed a novel deep learning algorithm which can be well generalized to new classes and the verification task. In [39], the authors proposed a novel method which reduce the inter-spectral differences significantly. The method effectively improves the matching between images taken different conditions. Lei et al. [40] proposed a method named discriminant face descriptor (DFD) which effectively enhances the discriminant ability of face representation.

## 3. Adaptive weighted fusion approach

### 3.1. Description of the approach

In this subsection, for simplicity of presentation, we assume that there are only two kinds of samples i.e. two kinds of data scores. The main steps of the adaptive weighted fusion approach (AWFA) are as follows.

Step 1. Feature extraction is performed for all samples including the test sample and the first and second kinds of samples.

Step 2. The scores of the test sample on the first kind of training samples (i.e. distances of the test sample to the first kind of training samples) are calculated and $d_i^1$ is used to stand for the distance of the test sample to the $i$-th class. $i=1,\ldots,C$ and $C$ is the total number of classes. The scores of the test sample on the second kind of training samples (i.e. distances of the test sample to the second kind of training samples) are calculated and $d_i^2$ is used to stand for the distance of the test sample to the $i$-th class. Define $\beta_r=\frac{h-\min(d_1^r,\ldots,d_C^r)}{h}, h=\sum_{j=1}^{C} d_j^r,\ r=1,2.$

Step 3. $d_i^1$ and $d_i^2$ are normalized to the range of 0 to 1 by using $d_i^1=(d_{\max}^1-d_i^1)/(d_{\max}^1-d_{\min}^1)$ and $d_i^2=(d_{\max}^2-d_i^2)/(d_{\max}^2-d_{\min}^2)$. $d_{\max}^1, d_{\min}^1$ are the maximum and minimum of $d_i^1$, respectively. $d_{\max}^2, d_{\min}^2$ are the maximum and minimum of $d_i^2$, respectively.

Step 4. $d_i^1$ are sorted in the order of ascending and the sorted result is recorded as $e_1^1\le e_2^1\le\ldots\le e_C^1$. $\hat{d}_1^2,\ldots,\hat{d}_C^2$ are sorted in the order of ascending and the sorted result is recorded as $e_1^2\le e_2^2\le\ldots\le e_C^2$. Let $w=(e_2^1-e_1^1)+(e_2^2-e_1^2)$, $w_1=(e_2^1-e_1^1)/w, w_2=(e_2^2-e_1^2)/w$. Because $e_1^2=e_1^1=0$, we also have $w_1=e_2^1/(e_2^1+e_2^2)$ and $w_2=e_2^2/(e_2^1+e_2^2)$. Let $f_i=\beta_1 w_1 d_i^1+\beta_2 w_2 d_i^2\quad(i=1,\ldots,C)$. If $k=\arg\min_i f_i$, then the test sample is assigned to the $k$-th class.

The algorithm of AWFA is described in Algorithm 1.

**Algorithm 1.**

1. Calculate $\beta_r=\frac{h-\min(d_1^r,\ldots,d_C^r)}{h},\quad h=\sum_{j=1}^{C} d_j^r,\ r=1,2.$
2. Calculate $d_i^1=(d_{\max}^1-d_i^1)/(d_{\max}^1-d_{\min}^1)$, $d_i^2=(d_{\max}^2-d_i^2)/(d_{\max}^2-d_{\min}^2)$.
3. Calculate $f_i=\beta_1 w_1 d_i^1+\beta_2 w_2 d_i^2$.
4. If $k=\arg\min_i f_i$, then the test sample is assigned to the $k$-th class.

### 3.2. Rationale and advantages of the proposed approach

The rationales and advantages of our approach (i.e. AWFA) are three-fold. First, it implements completely automatic weight selection. Second, it exploits the confidence (i.e. $\beta_g w_g,\ g=1,2$) of the scores in a very proper way. Third, differing from the majority of previous weighted score fusion approaches that assign fixed or adaptive weights to different kinds of scores (data sources) and do not take specific test sample into account, AWFA adaptively determine optimal weights for each test sample. This allows the dissimilarities between every test sample and each of the kinds of data sources to be elaborately and flexibly considered.

It is notable that in AWFA the setting of $w_1$ and $w_2$ are very reasonable owing to the following factor. $e_1^1$, $e_2^1$ are the best and secondary-best scores of the first kind of data sources. The study has demonstrated that a great difference between the best and secondary-best scores means reliable classification based on the scores [41,42]. In other words, it is proper to regard that the importance of classification results obtained using the first kind of data sources is proportional to the value of $e_2^1-e_1^1$. Similarly, we can think that the importance of classification results obtained using the second kind of data sources is proportional to the value of $e_2^2-e_1^2$. As a result, it is reasonable to take $w_1=(e_2^1-e_1^1)/w$ and $w_2=(e_2^2-e_1^2)/w$ as weights of the first and second kinds of data sources, respectively.

It should be pointed out that though Section 3.1 assumes that there are only two kinds of samples, i.e. two kinds of data sources, AWFA also can work for multiple kinds of data sources which is briefly described below. Suppose that there are $M$ kinds of data sources. For a test sample, let $t_i^1,\ldots,t_i^M$ be the distances (i.e. scores) of this test sample to the $i$-th class of the first to $M$-th kinds of data sources, respectively. In Step 4 of AWFA, after $e_2^1,\ldots,e_2^M$ are obtained, $M$ weights are generated using $w_r=\frac{e_2^r}{\sum_{j=1}^{M} e_2^j}\ (r=1,\ldots,M)$. $w_1,\ldots,w_M$ are respectively treated as the weights of the first to $M$-th kinds of data sources. Let $f_i=\sum_{j=1}^{M}\beta_j w_j d_i^j$ $(i=1,\ldots,C)$, $\beta_r=\frac{h-\min(d_1^r,\ldots,d_C^r)}{h}, h=\sum_{j=1}^{C} d_j^r$, $r=1,2,\ldots,M$. If $k=\arg\min_i f_i$, then the test sample is assigned to the $k$-th class. The above description shows that for multiple kinds of data sources, AWFA is also feasible.

It is notable that AWFA can directly work only under the condition that for every kind of data sources the distances (i.e. scores) of the test sample to each class are available. A typical case where this condition is not satisfied is that only the distances (i.e. scores) of the test sample to each training sample rather than to each class are available. In this case, in order to apply AWFA, we can use either of the following two schemes in advance. The first scheme is to exploit the distances (i.e. scores) of the test sample to each training sample to obtain the scores of the test sample to every class in advance and then to apply the naïve Steps 3 and 4 of AWFA. The second scheme is to modify Steps 3 and 4 of AWFA as follows.

---

Step 3. Let $s_i^1$ and $s_i^2$ denote distances (i.e. scores) of the test sample to the $i$-th training samples from the first and second kinds of data sources, respectively. $s_i^1$ and $s_i^2$ are normalized to the range of 0–1 by using $s_i^1=(s_{max}^1-s_i^1)/(s_{max}^1-s_{min}^1)$ and $s_i^2=(s_{max}^2-s_i^2)/(s_{max}^2-s_{min}^2)$. $s_{max}^1$, $s_{min}^1$ are the maximum and minimum of $s_i^1$, respectively. $s_{max}^2$, $s_{min}^2$ are the maximum and minimum of $s_i^2$, respectively.

Step 4. $s_i^1$ is sorted in the order of ascending and the sorted result is recorded as $g_1^1\le g_2^1\le\ldots\le g_N^1$. $N$ is the total number of the training samples from the first kind of data source. $s_i^2$ is sorted in the order of ascending and the sorted result is recorded as $g_1^2\le g_2^2\le\ldots\le g_N^2$. Let $w_1=g_2^1/(g_2^1+g_2^2)$ and $w_2=g_2^2/(g_2^1+g_2^2)$. Let $f_i=\beta_1 w_1 s_i^1+\beta_2 w_2 s_i^2\ (i=1,\ldots,N)$. If $k=\arg\min_i f_i$, then it is regarded that the test sample and the $k$-th training sample are from the same class.

---

## 4. Experimental results

In this section, we use five public datasets to test the performance of the proposed fusion approach. They are the Heterogeneous Face Biometrics (HFB) dataset [43], the 2D plus 3D palmprint dataset [44], the PolyU multispectral dataset [45], the Georgia Tech face dataset [46] and the Labeled Faces in the Wild (LFW) dataset [47]. To compare the performance of the proposed fusion approach, we choose Product rule [35], Sum rule [35], Min fusion [9], Max fusion [9] and "Average Score fusion" to conduct comparison experiments. "Average Score fusion" represents the score fusion approach in which the scores of the first and second kinds of data sources are first obtained using our approach and

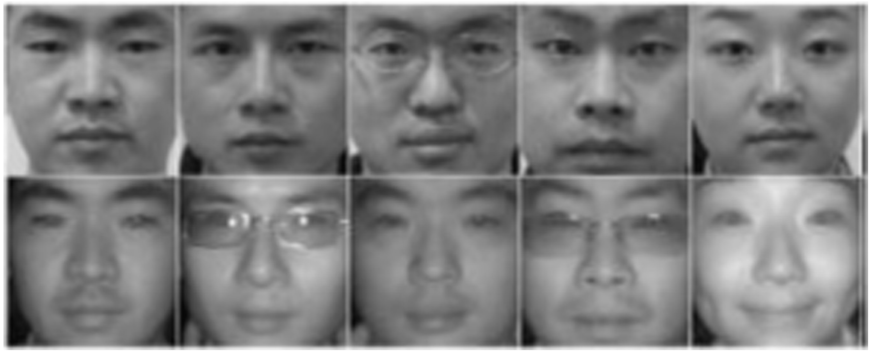

**Fig. 1.** Some visible and infrared face images. The first row shows the visible face images. The second row shows the infrared face images.

**Table 1**
Classification error rate (%) of either of NIR and VIS images from the HFB dataset.

| One training samples per subject | PCA | CRC | Original samples | SRC | LRC | Method in [39] | DFD |
|---|---|---|---|---|---|---|---|
| NIR | 7.35 | 5.33 | 10.33 | 4.72 | 5.56 | 4.67 | 2.67 |
| VIS | 12.00 | 10.00 | 17.00 | 8.89 | 10.55 | 7.00 | 5.67 |
| Two training samples per subject | PCA | CRC | Original samples | SRC | LRC | Method in [39] | DFD |
| NIR | 5.50 | 4.50 | 8.50 | 4.00 | 4.50 | 3.50 | 2.00 |
| VIS | 7.50 | 3.50 | 8.00 | 3.00 | 4.00 | 2.50 | 1.50 |

then they are combined with equal weights for classification. The score level fusion approaches presented in Section 2 are all based on the posteriori probabilities. For simplicity of implementation, in our experiments, when conducting comparison experiments we use only the Product rule and Sum rule presented in [35] and the Min rule and Max rule proposed in [7]. Based on these fusion schemes, we use principal component analysis (PCA) [48], collaborative representation based classification (CRC) [49], sparse representation based classifier (SRC) [50], linear regression classification (LRC) [51] as feature extraction procedure, respectively.

### 4.1. Experiments on the HFB

Heterogeneous Face Biometrics (HFB) [43] dataset contains 400 near infrared (NIR) images and 400 visual (VIS) images of 100 individuals, which are captured with various poses, expressions, light conditions and glasses accessories. For each pair of NIR and VIS images, they are not collected simultaneously and there are misalignments. Fig. 1 shows images of one subject in the HFB dataset.

Table 1 shows experimental results of either of NIR and VIS images. Table 2 shows experimental results of the proposed fusion approach and the conventional weighted matching score level fusion approaches. We can see that our approach outperforms the conventional weighted matching score level fusion. In all tables, the first row shows the feature extraction procedure. “Original samples” means that to directly use the original test and training samples to perform classification and there is no feature extraction procedure. Specifically, in Tables 1, 3, 5 and 8 (Tables 3, 5 and 8 are shown later), the classification error rate of “Original samples” is obtained by using the nearest neighbor classifier. In addition, for all experiments on PCA, the nearest neighbor classifier is also used to get the classification error rate.

From Table 1, we can see that DFD has the best classification accuracy among all the comparison methods. We take the NIR as an example. When we choose one training sample of each subject, the accuracy of DFD is 97.33%, which is 2% (=97.33–95.33%) higher than the closest competitor. From Table 2, AWFA has the best classification performance. When we choose one training sample of each subject, the accuracy of AWFA based PCA is 97.79%, which is 2.79% higher than the closest competitor.

**Table 2**
Classification error rate (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the HFB dataset.

| One training samples per subject | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 2.21 | 1.67 | 4.67 | 1.11 | 1.67 |
| Average score fusion | 5.27 | 1.67 | 6.00 | 1.39 | 1.94 |
| Sum rule | 5.56 | 1.94 | 6.39 | 1.67 | 2.22 |
| Product rule | 5.00 | 2.22 | 5.56 | 1.67 | 1.94 |
| Min rule | 5.56 | 2.78 | 6.94 | 1.94 | 2.50 |
| Max rule | 5.83 | 3.05 | 7.22 | 2.22 | 2.78 |
| Two training samples per subject | PCA | CRC | Original samples | SRC | LRC |
| AWFA | 1.50 | 0.50 | 3.00 | 0.50 | 1.00 |
| Average score fusion | 2.50 | 1.50 | 4.00 | 1.50 | 1.50 |
| Sum rule | 2.50 | 1.00 | 4.00 | 2.00 | 2.00 |
| Product rule | 3.00 | 1.00 | 4.50 | 1.50 | 1.50 |
| Min rule | 3.50 | 1.50 | 5.00 | 2.50 | 2.50 |
| Max rule | 3.50 | 1.50 | 5.00 | 2.50 | 2.50 |

**Table 3**
Classification error rate (%) of either of the 2D and 3D palmprint images.

| Images | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| 2D | 9.65 | 3.82 | 9.58 | 3.55 | 5.35 |
| 3D | 4.75 | 6.80 | 4.75 | 4.13 | 4.50 |

### 4.2. Experiments on the 2D plus 3D palmprint dataset

The 2D plus 3D palmprint dataset includes 8000 samples collected from 400 different palms [44]. Each palm has twenty 2D palmprint images and twenty 3D palmprint images. They were captured in two separated sessions and in every session ten samples of palmprint images of every palm were captured. It should be pointed out that here a sample consists of a 3D ROI (i.e. region of interest) and its corresponding 2D ROI. A 3D ROI is

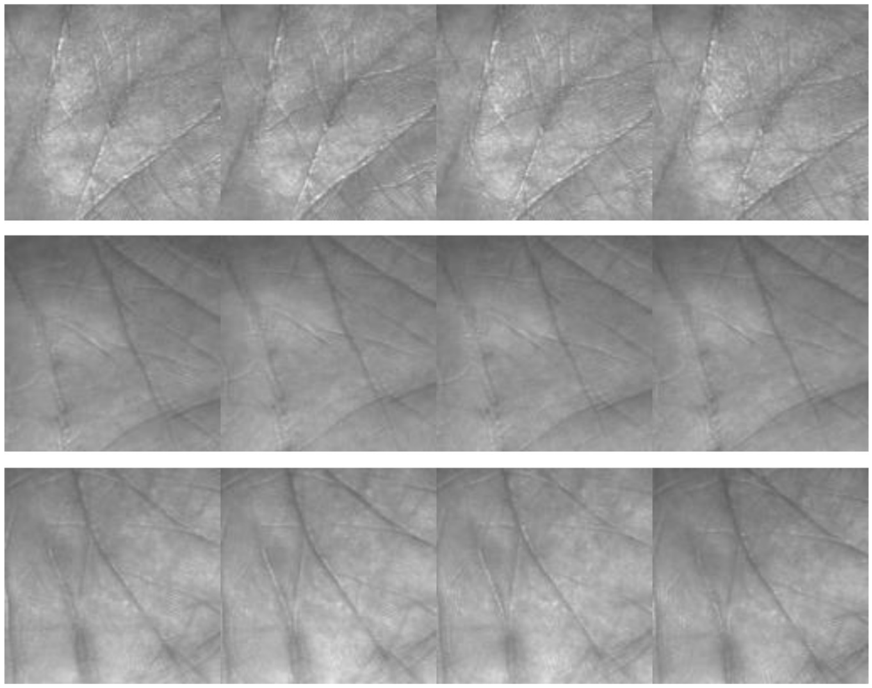

**Fig. 2.** Some 2D palmprint images from the 2D plus 3D palmprint dataset.

**Table 4**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the 2D plus 3D palmprint dataset.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 3.67 | 3.47 | 3.65 | 3.10 | 3.55 |
| Average score fusion | 4.23 | 3.90 | 3.50 | 3.45 | 4.10 |
| Sum rule | 4.67 | 4.10 | 3.68 | 3.15 | 3.98 |
| Product rule | 4.10 | 4.00 | 3.78 | 3.65 | 4.10 |
| Min rule | 5.20 | 4.55 | 4.05 | 4.13 | 4.78 |
| Max rule | 4.88 | 4.88 | 4.10 | 4.45 | 4.85 |

**Table 5**
Classification error rate (%) of blue, green or near infrared images of the Multispectral dataset.

| Channel | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| B | 5.67 | 3.27 | 4.80 | 3.46 | 4.10 |
| G | 5.62 | 3.43 | 7.90 | 3.16 | 3.67 |
| I | 3.57 | 3.70 | 5.23 | 3.50 | 3.96 |

represented by a binary file composed of $128 \times 128$ mean curve ratios, and every 2D ROI is represented by a BMP format image file. We used the first four samples (i.e. four 2D ROI images and four 3D ROI images) collected in the first session as training samples and regard all 10 samples collected in the second session as test samples. We resized every image to a 32 by 32 matrix and converted it into a 1024-dimensional unit vector with length of 1 before they were used to test the approaches. Fig. 2 shows some 2D palmprint images from the 2D plus 3D palmprint dataset.

Table 3 shows the experimental results of either of 2D and 3D images. Table 4 shows the experimental results of the proposed fusion approach and the conventional weighted matching score level fusion approaches. We can see again that our approach is superior to the conventional weighted matching score level fusion in the classification error rate.

**Table 6**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the Multispectral dataset with the score fusion of blue and near infrared images.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 3.26 | 3.10 | 4.97 | 2.96 | 3.15 |
| Average score fusion | 4.10 | 3.20 | 5.10 | 3.46 | 3.98 |
| Sum rule | 4.02 | 3.15 | 5.17 | 3.33 | 4.10 |
| Product rule | 3.95 | 3.50 | 5.23 | 3.67 | 4.23 |
| Min rule | 4.56 | 3.87 | 5.68 | 4.12 | 4.98 |
| Max rule | 4.87 | 4.12 | 5.79 | 4.65 | 5.12 |

### *4.3. Experiments on the PolyU multispectral palmprint dataset*

The PolyU multispectral palmprint dataset was collected from 250 subjects (55 women and 195 men) using the palmprint acquisition device developed by PolyU [45]. Each subject provided palmprint images of both the left and right palms. There were four illuminations, i.e. red, green, blue and near infrared illuminations, so there were four kinds of palmprint images, i.e. red, green, blue, and near infrared palmprint images. These multispectral palmprint images were collected in two separate sessions. In each session, every palm provided 6 palmprint images at each spectral band. As a result, for each spectral band, the dataset contained 6000 images from 500 different palms. In the following experiments, we use the first three images of each spectral band of a palm from the first session as training samples and exploit all six images of each spectral band of a palm from the second session as testing samples. The resolution of the palmprint image is $352 \times 288$. The $128 \times 128$ region of interest (ROI) domain is extracted from each palmprint image using the approach proposed in [52]. The ROI images are resized to $32 \times 32$ images . Fig. 3 shows some palmprint images from the PolyU multispectral palmprint dataset.

Table 5 shows classification error rates of the blue, green or near infrared images of the Multispectral dataset, respectively. Table 6 shows the classification error rates of the proposed fusion

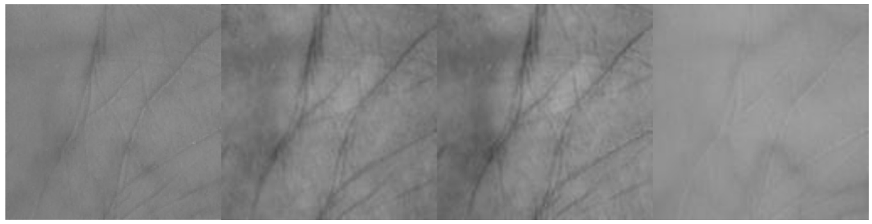

**Fig. 3.** Four ROI images of a same palm. The first, second, third and fourth ROI images were extracted from the red, green, blue and near infrared images, respectively.

**Table 7**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the Multispectral dataset with the score fusion of green and near infrared images.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 2.63 | 1.10 | 2.11 | 0.96 | 1.68 |
| Average score fusion | 4.11 | 1.23 | 2.33 | 1.26 | 2.10 |
| Sumrule | 4.01 | 1.20 | 2.47 | 1.35 | 2.22 |
| Product rule | 4.26 | 1.33 | 2.46 | 1.66 | 2.64 |
| Min rule | 4.53 | 1.84 | 2.57 | 1.87 | 2.18 |
| Max rule | 4.66 | 1.76 | 2.68 | 2.31 | 2.68 |

**Fig. 4.** Some image samples from the GT face dataset.

**Table 8**
Classification error rate (%) of the R, G, B color channels of the GT face dataset.

| | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| R-channel | 44.00 | 44.40 | 48.40 | 38.20 | 38.20 |
| G-channel | 50.40 | 46.80 | 52.20 | 40.20 | 41.40 |
| B-channel | 54.80 | 50.20 | 56.40 | 43.20 | 46.40 |

approach and the conventional weighted matching score level fusion approaches on the Multispectral dataset with the score fusion of blue and near infrared images. Table 7 shows the classification error rates of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the Multispectral dataset with the score fusion of green and near infrared images. From the experimental results, we can see that AWFA performs better than the conventional weighted matching score level fusion approaches.

### 4.4. Experiments on the Georgia Tech face dataset

The Georgia Tech (GT) face dataset was built at Georgia Institute of Technology. This dataset contains images of 50 people taken in two or three sessions. Each subject in the dataset is represented by 15 color JPEG images with cluttered background taken at resolution $640 \times 480$ pixels. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions and scales. Each image was manually labeled to determine the position of the face in the image. In our experiments, we resize all images to $60 \times 50$ pixels. Fig. 4 shows some image samples from the GT face dataset. The first five images of each subject are used as training samples and the rest samples are taken as test samples.

Table 8 shows classification error rates of the R, G, B color channels, respectively. Tables 9–11 show the classification error rates of the proposed fusion approach and the conventional weighted matching score level fusion approaches based on the score fusion of two color channels from the original three color channels, respectively. From the experimental results, we can see that AWFA performs better than the conventional weighted matching score level fusion approaches.

### 4.5. Experiments on the LFW dataset

Labeled Faces in the Wild (LFW) [47] is a database collected from the web for studying the problem of unconstrained face recognition. There are 13,233 images from 5749 different people, with large pose, occlusion, expression variations. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the dataset. The only constraint on these faces is that they were detected by the Viola–Jones face detector. In our experiments, we choose 1251 images from 86 people images [47]. Each image was manually

**Table 9**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the GT face dataset based on the score fusion of the red and green channels.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 40.20 | 39.60 | 46.60 | 35.20 | 36.60 |
| Average score fusion | 43.00 | 41.20 | 47.20 | 36.20 | 37.20 |
| Sum rule | 43.20 | 40.80 | 47.60 | 36.40 | 37.80 |
| Product rule | 42.80 | 41.60 | 47.20 | 36.60 | 36.80 |
| Min rule | 43.60 | 42.20 | 47.60 | 37.20 | 37.20 |
| Max rule | 43.80 | 42.60 | 47.80 | 37.80 | 38.00 |

**Table 10**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the GT face dataset based on the score fusion of the blue and green channels.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 46.20 | 42.20 | 48.60 | 37.20 | 38.20 |
| Average score fusion | 47.80 | 43.80 | 49.20 | 38.60 | 39.60 |
| Sum rule | 47.40 | 43.60 | 49.60 | 38.40 | 39.40 |
| Product rule | 47.60 | 44.20 | 50.00 | 39.00 | 39.00 |
| Min rule | 48.20 | 44.60 | 50.20 | 39.20 | 40.20 |
| Max rule | 48.00 | 44.80 | 50.60 | 39.40 | 40.60 |

**Table 11**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the GT face dataset based on the score fusion of the red and blue channels.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 42.00 | 42.20 | 45.20 | 36.00 | 36.20 |
| Average score fusion | 43.20 | 43.20 | 46.40 | 37.20 | 37.60 |
| Sum rule | 43.60 | 43.00 | 46.20 | 37.40 | 37.80 |
| Product rule | 44.00 | 43.60 | 46.00 | 37.00 | 37.40 |
| Min rule | 43.60 | 43.80 | 47.20 | 37.80 | 38.00 |
| Max rule | 44.20 | 44.00 | 47.60 | 38.00 | 38.20 |

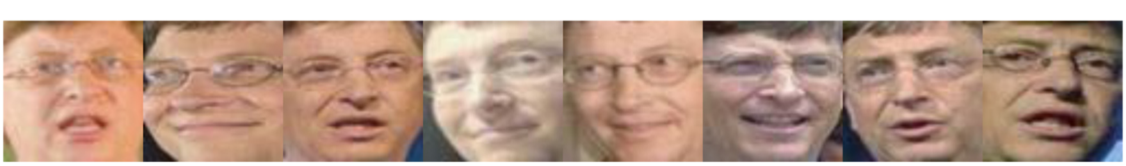

**Fig. 5.** Sample images of one individual from the LFW dataset.

**Table 12**
Classification error rate (%) of the R, G, B color channels of the LFW face dataset.

| | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| R-channel | 74.07 | 64.48 | 75.49 | 61.28 | 65.19 |
| G-channel | 75.49 | 66.79 | 76.20 | 62.52 | 65.36 |
| B-channel | 74.78 | 65.19 | 76.38 | 62.70 | 65.72 |

**Table 13**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the LFW face dataset based on the score fusion of the red and green channels.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 66.25 | 63.59 | 66.61 | 55.24 | 56.66 |
| Average score fusion | 66.96 | 65.19 | 68.21 | 57.19 | 57.19 |
| Sum rule | 68.21 | 65.72 | 67.67 | 57.37 | 57.73 |
| Product rule | 68.74 | 65.54 | 67.14 | 57.55 | 58.79 |
| Min rule | 70.52 | 67.14 | 69.63 | 59.15 | 59.15 |
| Max rule | 69.80 | 68.56 | 70.87 | 58.79 | 60.00 |

**Table 14**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the LFW face dataset based on the score fusion of the blue and green channels.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 66.25 | 64.12 | 68.56 | 57.19 | 59.15 |
| Average score fusion | 67.85 | 65.72 | 69.27 | 58.61 | 61.63 |
| Sum rule | 67.32 | 65.54 | 69.63 | 58.44 | 61.46 |
| Product rule | 67.67 | 66.25 | 69.98 | 58.97 | 62.70 |
| Min rule | 68.20 | 67.67 | 71.23 | 59.15 | 63.23 |
| Max rule | 68.03 | 68.74 | 71.58 | 59.33 | 63.59 |

**Table 15**
Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the LFW face dataset based on the score fusion of the red and blue channels.

| Fusion approaches | PCA | CRC | Original samples | SRC | LRC |
|---|---|---|---|---|---|
| AWFA | 63.94 | 62.17 | 65.19 | 55.95 | 59.15 |
| Average score fusion | 65.19 | 63.23 | 66.43 | 57.19 | 60.57 |
| Sum rule | 65.54 | 63.59 | 66.25 | 57.37 | 61.81 |
| Product rule | 64.12 | 63.77 | 66.07 | 57.02 | 61.46 |
| Min rule | 66.61 | 64.83 | 67.14 | 57.73 | 61.99 |
| Max rule | 66.25 | 65.01 | 67.67 | 58.08 | 62.17 |

cropped and resized to $40 \times 50$ pixels. The sample images of one individual from the LFW database are showed in Fig. 5. The first eight images of each subject as training samples and the rest samples are taken as test samples

Table 12 shows classification error rate of the comparison methods on the R, G, B color channels of the LFW face dataset, respectively. Tables 13–15 show the classification error rate of the comparison fusion methods with the score fusion of two color channels from the three color channel, respectively. From the experim conventional weighted matching score level fusion methods again.ental results, we can see that AWFA outperforms the

Table 12 shows classification error rates of the R, G, B color channels, respectively. Tables 13–15 show the classification error rates of the proposed fusion approach and the conventional weighted matching score level fusion approaches based on the score fusion of two color channels from the original three color channels, respectively. From the experimental results, we can see that AWFA performs better than the conventional weighted matching score level fusion approaches again.

## 5. Conclusions

The adaptive weighted fusion approach proposed in this paper can automatically set optimal weights for every test sample and it does not need any manual setting. As a result, this approach can well integrate the advantages of the complementary data sources. The extensive experiments demonstrate that the devised approach outperforms previous state-of-the-art score fusion approaches.

## References


[1] K. McDonald, A.F. Smeaton, A Comparison of Score, Rank and Probability-Based Fusion Methods for Video Shot Retrieval, CIVR, 2005, pp. 61–70.
[2] A. Nagar, K. Nandakumar, A.K. Jain, Multibiometric cryptosystems based on feature-level fusion, IEEE Trans. Inf. Forensics Secur. 7 (1) (2012) 255–268.
[3] K. Nandakumar, A.K. Jain, A. Ross, Fusion in Multibiometric Identification Systems: What about the Missing Data? ICB, 2009, pp. 743–752.
[4] K. Nandakumar, Y. Chen, S.C. Dass, A.K. Jain, Likelihood ratio-based biometric score fusion, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2008) 342–347.
[5] A. Ross, K. Nandakumar, Fusion, score-level, Encyclopedia of Biometrics, 2009, pp. 611–616.
[6] K.C. Sim, K.-A. Lee, Adaptive score fusion using Weighted Logistic Linear Regression for spoken language recognition, ICASSP, 2010, pp. 5018–5021.
[7] D. Zhang, F.X. Song, Y. Xu, Z.Z. Liang, Advanced Pattern Recognition Technologies with Applications to Biometrics, Medical Information Science Reference, New York, 2009.
[8] Y. Xu, D. Zhang, Represent and fuse bimodal biometric images at the feature level: complex-matrix-based fusion scheme, Opt. Eng. 49 (3) (2010) 037002.
[9] A.K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognit. 38 (12) (2005) 2270–2285.
[10] Y. Xu, Q. Zhu, D. Zhang, J.-Y. Yang, Combine crossing matching scores with conventional matching scores for bimodal biometrics and face and palmprint recognition experiments, Neurocomputing 74 (2011) 3946–3952.
[11] R. Brunelli, D. Falavigna, Person Identification Using Multiple Cues, IEEE Trans. Pattern Anal. Mach. Intell. 17 (10) (1995) 955–966.
[12] Y. Ma, B. Cukic, and H. Singh, A classification approach to multibiometric score fusion, in: Proceedings of the fifth International Conference on Audio Video-Based Biometric Person Authentication, July 2005, pp. 484–493.
[13] S. Bengio, C. Marcel, S. Marcel, J. Mariethoz, Confidence measures for multimodal identity verification, Inf.Fusion 3 (4) (2002) 267–276.
[14] M.S. Moin, M. Parviz, Exploring AUC Boosting Approach in Multimodal Biometrics Score Level Fusion IIH-MSP, IEEE Computer Society, Kyoto, Japan (2009) 616–619.
[15] M. Parviz, and M.S. Moin, Multivariate polynomials estimation based on gradientboost in multimodal biometrics, in: Proceedings of the Conference on ICIC, vol. 15 (3) , 2008, pp. 471–477.
[16] R. Snelick, U. Uludag, A. Mink, M. Indovina, A.K. Jain, Large scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 450–455.
[17] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas., On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) 226-239 20 (3) (1998).
[18] B. Ulery, A.R. Hicklin, C. Watson, W. Fellner, P. Hallinan, Studies of Biometric Fusion Technical Report IR 7346, NIST, Gaithersburg, MD, 2006.
[19] A. Rakhlin, D. Panchenko, S. Mukherjee, Risk bounds for mixture density estimation, ESAIM: Probab. Stat. 9 (2005) 220–229.
[20] K.A. Toh, J. Kim, S. Lee, Maximizing area under roc curve for biometric scores fusion, Pattern Recognit. 41 (11) (2008) 3373–3392.
[21] C. Rudin, R.E. Schapire, Margin-based ranking and an equivalence between adaboost and rankboost, Journal of Machine Learning Research 10 (2009) 2193–2232.
[22] S. Gao, Q. Sun, Improving semantic concept detection through optimizing ranking function, IEEE Trans. Multimed. 9 (7) (2007) 1430–1442.
[23] H.M. Sim, H. Asmuni, R. Hassan, R.M. Othman, Multimodal biometrics: weighted score level fusion based on non-ideal iris and face images, Expert Syst. Appl. 41 (11) (2014) 5390–5404.
[24] S. McCloskey, J. Liu, Metadata-Weighted Score Fusion for Multimedia Event Detection, CRV, 2014, pp. 299–305.
[25] C. Butakoff, A.F. Frangi, A Framework for Weighted Fusion of Multiple Statistical Models of Shape and Appearance, IEEE Trans. Pattern Anal. Mach. Intell. 28 (11) (2006) 1847–1857.
[26] Y. Xu, X. Li, J. Yang, Z. Lai, D. Zhang, Integrating conventional and inverse representation for face recognition, IEEE Trans. Cybern. 44 (10) (2013) 1738–1746. http://dx.doi.org/10.1109/TCYB.2013.2293391.
[27] P.G. Casazza, J. Peterson, Weighted fusion frame construction via spectral tetris, Adv. Comput. Math. 40 (2) (2014) 335–351.
[28] A.K. Jain, A.A. Rosss, Learning user-specific parameters in a multibiometric system, in: Proceedings of the International Conference on Image Processing (ICIP), Rochester, New York, September 22–25, 2002.
[29] M.E. Moumene, R. Nourine, D. Ziou, Generalized Exposure Fusion Weights Estimation, pp. 71–76.
[30] S. Kim, J.Y. Choi, S. Han, Y.M. Ro, Adaptive weighted fusion with new spatial and temporal fingerprints for improved video copy detections, Signal Process.: Image Commun. 29 (7) (2014) 788–806.
[31] G. Gao, L. Zhang, J. Yang, L. Zhang, D. Zhang, Reconstruction based finger-knuckle-print verification with score level adaptive binary fusion, IEEE Trans. Image Process. 22 (12) (2013) 5050–5062.
[32] S. Yang, W. Zuo, L. Liu, Y. Li, D. Zhang, Adaptive Weighted Fusion of Local Kernel Classifiers for Effective Pattern Classification, ICIC, (1), 2011, pp. 63–70.
[33] S. Prabhakar, A.K. Jain, Decision-level fusion in fingerprint verification, Pattern Recognit. 35 (4) (2002) 861–874.
[34] E.S. Bigun, J. Bigun, B. Duc, S. Fischer, Expert conciliation for multimodal person authentication systems using Bayesian statistics, in: Proceedings of the first International Conference on AVBPA, Crans-Montana, Switzerland, 1997, pp. 291–300.
[35] A. Ross, A.K. Jain, Information fusion in biometrics, Pattern Recognit. Lett. 24 (13) (2003) 2115–2125.
[36] R. Snelick, M. Indovina, J. Yen, A. Mink, Multimodal biometrics: issues in design and testing, in: Proceedings of the fifth International Conference on Multimodal Interfaces, Vancouver, Canada, 2003, pp. 68–72.
[37] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, DeepFace: closing the gap to human-level performance in face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1701–1708.
[38] Y. Sun, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[39] J. Chen, D. Yi, J. Yang, G. Zhao, S.Z. Li, and M. Pietikainen, Learning mappings for face synthesis from near-infrared to visual light images, in: Proceeings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 156–163.
[40] Z. Lei, M. Pietikainen, S.Z. Li, Learning discriminant face descriptor, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 289–302.
[41] C. Kim, C.-H. Choi, Image covariance-based subspace method for face recognition, Pattern Recognit. 40 (5) (2007) 1592–1604.
[42] Jeffery R. Price, Timothy F. Gee, Face recognition using direct weighted linear discriminant analysis and modular subspaces, Pattern Recognit. 38 (2) (2005) 209–219.
[43] S.Z. Li, Z. Lei, M. Ao, The HFB face dataset for heterogeneous face biometrics research, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2009, pp. 1–8.
[44] HK-PolyU 2D+3D palmprint dataset: ⟨http://www.comp.polyu.edu.hk/~biometrics/2D_3D_Palmprint.htm⟩.
[45] D. Han, Z. Guo, D. Zhang, Multispectral palmprint recognition using wavelet based image fusion, in: Proceedings of the International Conference of the Signal Process, 2007, pp. 1–6.
[46] S. Wang, J. Yang, N. Zhang, C.-G. Zhou, Tensor discriminant color space for face recognition, IEEE Trans. Image Process. 20 (9) (2011) 2490–2501.
[47] S. Wang, J. Yang, M.F. Sun, X.J. Peng, M.M. Sun, C.G. Zhou, Sparse tensor discriminant color space for face verification, IEEE Trans. Neural Netw. Learn. Syst. 23 (6) (2012) 876–888.
[48] P. Belhumeur, J. Hepanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[49] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 6669, (5), 2011, pp. 471–478.
[50] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[51] I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106–2112.
[52] D. Zhang, Z. Guo, G. Lu, L. Zhang, W. Zuo, An online system of multispectral palmprint verification, IEEE Trans. Instrum. Meas. 59 (2) (2010) 480–490.


**Yong Xu** (M'06) was born in Sichuan, China, in 1972. He received the B.S. and M.S. degrees in 1994 and 1997, respectively, and the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, Nanjing, China, in 2005. Currently, he is with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China. His current research interests include pattern recognition, biometrics, machine learning, image processing, and video analysis.

**Yuwu Lu** received the B.S. and M.S. degrees in 2008 and 2011, respectively. He is currently pursuing the Ph. D. degree in computer science and technology at Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China.

He has published 8 scientific papers. His current research interests include pattern recognition and machine learning.