Multi-classifier ensemble based on dynamic weights

Fuji Ren [1,2] • Yanqiu Li [1] • Min Hu [1]

Abstract In this study, a novel multi-classifier ensemble method based on dynamic weights is proposed to reduce the interference of unreliable decision information and improve the accuracy of fusion decision. The algorithm defines decision credibility to describe the real-time importance of the classifier to the current target, combines this credibility with the reliability calculated by the classifier on the training data set and dynamically assigns the fusion weight to the classifier. Compared with other methods, the contribution of different classifiers to fusion decision in acquiring weights is fully evaluated in consideration of the capability of the classifier to not only identify different sample regions but also output decision information when identifying specific targets. Experimental results on public face databases show that the proposed method can obtain higher classification accuracy than that of single classifier and some popular fusion algorithms. The feasibility and effectiveness of the proposed method are verified.

1 Introduction

The advent of the information age has increased the amount of data from scientific research in various fields of social life. Realising massive data storage and intelligent processing and fully utilising the knowledge and value contained in data have been the main tasks of the academia and the industry [5, 16, 30, 56–58]. Machine learning is a mainstream intelligent data processing technology and is an important way to achieve the above-mentioned goals. The continuous

---

✉ Yanqiu Li
liyanqiu2012@163.com

[1] School of Computer and Information, Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, Hefei University of Technology, Hefei 230009, China

[2] Graduate School of Advanced Technology&Science, University of Tokushima, Tokushima 7708502, Japan

development of science and technology has facilitated the development of machine learning and its application to various fields, such as 3D modeling [55] and forecasting and analysis of stock trend [52, 54, 59]. Classification technology is an indispensable part of machine learning. Machine learning solves classification problems by finding a fully good function, that is, the classifier, by use of a specific learning algorithm in the hypothetical space to simulate the actual classification function. The widely used single classifier models are forward feedback artificial (BP) neural network [40], support vector machine (SVM) [10] and decision tree [27]. BP neural network is a non-linear mapping system with a forward structure without a feedback one. This model can fit the input–output relationship of nonlinear systems effectively and have been used by many scholars in this field. SVM is a supervised learning model. This model can transform linear inseparable samples into high-dimensional space by use of a nonlinear mapping algorithm to make the samples linearly separable and can effectively deal with binary classification problems. Unlike the two previous models, the construction of the decision tree does not depend on specific domain knowledge. This model uses the attribute selection metric to determine the topology amongst various feature attributes and thus complete the classification decision. Although each of these classifier models has achieved good results, all of them present poor generalisation capability and of classification accuracy that cannot meet users' requirements when noise exists in the input sample or the distribution of the training and classification sets is inconsistent. Ensemble learning is a new paradigm of machine learning. This model integrates multiple classifiers in accordance with specific integration rules to solve the classification problem and can thus improve the accuracy and generalisation capability of system prediction. Ensemble learning is widely used in various fields, such as human activity recognition [33–35, 39], human motion tracking [11, 32], prediction [36, 44], water quality supervision [37, 38] and face or emotion discrimination [4, 14, 20].

Multi-classification ensemble can be divided into static ensemble and dynamic ensemble depending on whether weights are adaptively assigned to each classifier when different query samples are identified. The classifier that participates in the static integration is no longer changed after its determination. Castrillón-Santana et al. [47] focused on gender classification based on face images and evaluated multi-expert systems using three different fusion protocols. However, the said methods ignore the importance of classifiers when identifying different query samples and assign the same weight value to each expert. Adaboost [43] is a widely used static weighted ensemble algorithm. The weights of the base classifiers are evaluated before the query samples arrive, and these base classifiers are integrated by linear weighting to obtain the final decision. Kuncheva [25] proposed a method based on probability weighted voting. Chen [7] proposed a method based on training error to calculate the weight coefficient in completing the SAR image classification. The aforementioned static methods enable the classification model to statistically achieve the best outcome; however, a corresponding change cannot be made with the change in the input information and the abnormal output of the classifier. Such condition is obviously unreasonable. By contrast, dynamic methods can determine the adaptive weights on the basis of the characteristics of the sample; these weights can effectively emphasise the decision-making contribution made by excellent classifiers and suppress the influence of unreliable information output by use of high-deviation classifiers on classification. Cheon [8] assigned weights following the Hausdorff distance between query and gallery image sequences. Zhang [60] considered the performance of each classifier in different sample environments and then proposed a dynamic adaptive weighted voting method. Considering the correlation and reliability of the query descriptors, Sun [49] calculated the weights on the basis of the sparse coefficients of the query sample in different classes. The results showed that the weights change

with the change in the number of query sample. Most of the existing weight allocation methods assign weights to the classifier from a certain aspect, and none of the methods either based on the statistical performance of the classifier or the characteristic distribution of the query sample comprehensively consider the identification performance of the classifier for different sample regions and samples. Each classifier presents different recognition capabilities for different samples owing to the different characteristics of these samples, and the recognition performance of the same classifier in different regions of the sample space also differs. Therefore, we determine whether the weight of the classifier in the fusion process can be dynamically adjusted by estimating the recognition performance of the classifier on different sample regions and samples, thereby reducing the interference of unreliable information on the decision making and achieving good ensemble performance. This study provides the following contributions.

- A dynamic weighted multi-classifier ensemble method that defines reliability and credibility is proposed. Weights can be dynamically changed with the query samples, and the importance of the classifier in the fusion process can be described fully.
- The reliability of the classifier is defined on the basis of its recognition capability, which is determined by the prior knowledge gained during the training process. Fuzzy set theory and cloud model are utilised to evaluate this criterion for base classifiers.
- The credibility of the classifier is obtained by calculating the posterior probability distribution, which is used to determine the separability feature of the classifier. A mapping function is used to map the separability into the credibility of the classifier.

Compared with single classifier and current popular fusion methods, our method can more effectively reduce the effect of unreliable instance information in the training phase of the classifier and can therefore fuse the decisions in classification efficiently and improve the overall performance of the integrated method.

The rest of the paper is organised as follows. Section 2 reviews the traditional multi-classifier ensemble method. Section 3 presents the proposed method of multi-classifier ensemble based on dynamic weights. Section 4 demonstrates the application of the proposed algorithm in face recognition. Section 5 provides the experimental results and analysis. Section 6 contains the discussion. Finally, Section 7 elaborates the conclusions of the study.

2 Related work

On the basis of the output of the classifier, the decision information can be divided into three levels [53]: abstract, ranked and measurement levels. At the abstract level, each classifier directly outputs the category labels. Common ensemble methods include voting [42], weighted voting and behavioural knowledge space. The ranked level sorts the output results of the classifier. This type of output usually occurs in problems with a large number of categories. The commonly used ensemble methods of ranked level are highest rank method, Borda count method [19] and logistic regression. At the measurement level, the output of each classifier is the probability, credibility or distance measure of the class. The measurement level is more informative than the abstract and ranked levels. The measurement level also contains a richer amount of decision information compared with that of the two other levels. The corresponding ensemble methods include Max/ Min/ Sum, linear combination method, D-S evidence theory [17] and fuzzy integral method [9, 26]. The multi-classifier ensemble method based on measurement level is defined as follows.

We define a classification space S, which contains M different classes, for K classifiers. Furthermore, $S = C_1 \cup \cdots \cup C_M$, C_m indicates the m-th class, $1 \leq m \leq M$, $E = \{e_1, e_2, \cdots e_K\}$ is a collection of base classifiers and $p_{km}(x)$ represents the posterior probability of query sample x discriminated by classifier e_k as belonging to class m. Thus, the subsequent decision output matrix can be obtained by Eq. (1) as follows:

$$\begin{bmatrix} p_{11}(x) & p_{12}(x) & \cdots & p_{1M}(x) \\ p_{21}(x) & p_{22}(x) & \cdots & p_{2M}(x) \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1}(x) & p_{K2}(x) & \cdots & p_{KM}(x) \end{bmatrix}. \tag{1}$$

The multi-classifier ensemble obtains the final classification decision by merging the recognition outputs of each classifier in different categories. The decision output of the multi-classifier system on class C_m is obtained by Eq. (2) as follows:

$$P_m(x) = \sum_{k=1}^{K} p_{km}(x). \tag{2}$$

We select the category with the maximum posterior probability sum as the class to which query sample x belongs and complete the final decision. In particular, the decision result of the joint classification is as follows:

$$E(x) = \max_{j} P_j(x), 1 \leq j \leq M. \tag{3}$$

To highlight the importance of different classifiers in the fusion process, Eq. (2) is modified as follows:

$$P_m(x) = \sum_{k=1}^{K} \omega_k p_{km}(x). \tag{4}$$

In Eq. (4), ω_k represents the weight of classifier e_k. High value of ω_k means high importance of the classifier in the fusion process.

3 Multi-classifier ensemble based on dynamic weights

In this study, we propose a multi-classifier weighted ensemble method based on measurement level fusion. Figure 1 presents the framework of our approach that comprises three major components. (1) In the training phase, the fuzzy density of the classifier is computed firstly and then inputted into an inverse cloud generator in accordance with the correct and incorrect decisions of the corresponding instances. As a result, the reliability of the classifier can be modelled on the basis of the expected value, entropy and hyper entropy digital eigenvalues. (2) The separability of the classifier is obtained by calculating the posterior probability distribution when a specific target is identified. Thereafter, a mapping function is used to map the separability into the credibility of the classifier. (3) The dynamic weights of the classifier are obtained on the basis of reliability and decision credibility. We utilise the dynamic weights to implement dynamic weighted integration. The three major components are discussed in detail in the following sections.
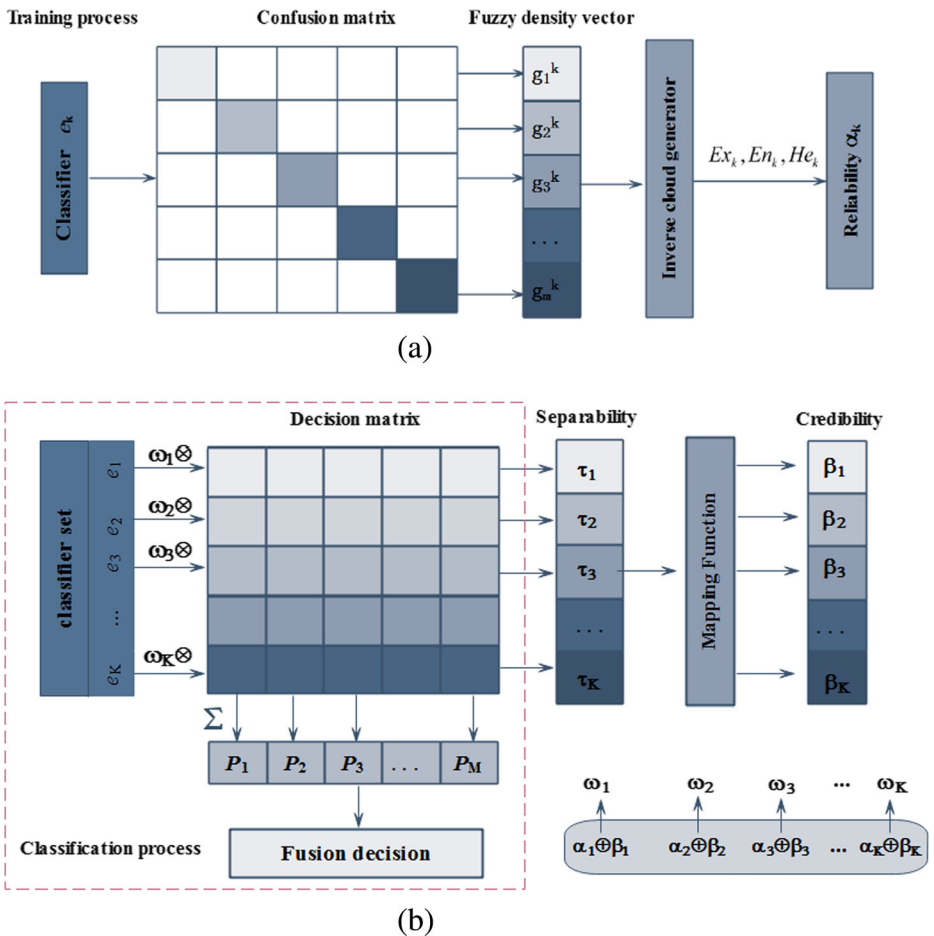
Fig. 1 Framework of multi-classifier ensemble based on dynamic weight: (a) acquisition process of classifier reliability; (b) diagram for obtaining classifier credibility and fusion decision

3.1 Statistic of fuzzy density

We aim to obtain the reliability of the classifier. For this purpose, we use different learning algorithms firstly to train the data set for constructing the base classifier and then use confusion matrix $CM^k_{M \times M}$ to compute the identification of the known class samples made by base classifier e_k. The expression is as follows:

$$CM^k_{M \times M} = \begin{bmatrix} n^k_{11} & n^k_{12} & \cdots & n^k_{1M} \\ n^k_{21} & n^k_{22} & \cdots & n^k_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ n^k_{M1} & n^k_{M2} & \cdots & n^k_{MM} \end{bmatrix}, \quad (5)$$

where $k = 1, 2, \cdots, K$ and n^k_{ml} is the number of samples in class C_m identified by base classifier e_k as class C_l. If $m = l$, then n^k_{ml} is the number of samples in class C_m correctly identified by

base classifier e_k; if $m \neq l$, then n_{ml}^k is the number of samples in class C_m incorrectly identified by base classifier e_k as class C_l.

For K different base classifiers, we can obtain K confusion matrixes CM^1, CM^2, $\cdots$, CM^K to determine the recognition capability of classifier $e_k(k=1, 2, \cdots, K)$ to $C_m(1 \leq m \leq M)$. We use fuzzy integral theory to compute the fuzzy density of the classifier by confusion matrix CM^k. Fuzzy density can be calculated by

$$g_m^k = \left[\frac{1}{M-1} \sum_{n=1, n \neq m}^{M} \left(1 - \frac{n_{nm}^k}{\sum_{l=1}^{M} n_{nl}^k} \right) \right] \frac{n_{mm}^k}{\sum_{l=1}^{M} n_{ml}^k}. \tag{6}$$

In Eq. (6), $\dfrac{n_{mm}^k}{\sum_{l=1}^{M} n_{ml}^k}$ is the ratio that class C_m is correctly recognised by classifier e_k; $\dfrac{n_{nm}^k}{\sum_{l=1}^{M} n_{nl}^k}$ is the ratio that class C_n is incorrectly recognised by classifier e_k as class C_m, $n \neq m$. Large value of $\dfrac{n_{mm}^k}{\sum_{l=1}^{M} n_{ml}^k}$ means small value of $\dfrac{n_{nm}^k}{\sum_{l=1}^{M} n_{nl}^k}$ and large value of g_m^k, thereby indicating that the recognition capability of classifier e_k to class C_m is strong.

The recognition capability of classifier e_k to identify M classes can be represented by fuzzy density vector $G^k = \left(g_1^k, g_2^k, \cdots, g_M^k \right)$. K different base classifiers make up fuzzy density matrix GM as follows:

$$GM = \left[G^1, G^2, \cdots, G^K \right]^T = \begin{bmatrix} g_1^1 & g_2^1 & \cdots & g_M^1 \\ g_1^2 & g_2^2 & \cdots & g_M^2 \\ \vdots & \vdots & \ddots & \vdots \\ g_1^K & g_2^K & \cdots & g_M^K \end{bmatrix}. \tag{7}$$

3.2 Basic concept and eigenvalue acquisition of cloud model

The cloud model is introduced to describe the reliability of the identification results made by the classifier in different sample regions. The cloud model is a type of uncertain transformation model between qualitative concept and quantitative data proposed by Li [29]. The model combines the fuzziness in fuzzy set theory with the randomness in probability theory to characterise a whole concept. The model also occupies an important position in performance evaluation, recognition and other areas [12, 31, 48]. The cloud model consists of three eigenvalues: expected value Ex, entropy En and hyper entropy He. Ex is the central value of qualitative language concept domain and reflects the situation of gravity of cloud drop group. En is a measure of the ambiguity of the qualitative concept. On the one hand, it reflects the range of the concept that can be accepted in the domain space (i.e. the ambiguity). On the other hand, the points reflected in the domain space can represent the probability of this concept, thereby indicating the randomness of cloud droplets of the qualitative concept. He represents the aggregation of cloud droplets in the domain space and reflects the discrepancy in the cloud droplets and the random change in membership. In the number field space, He represents the cohesion of the uncertainty of all points of the language value. The cloud generator can be in

forward and reverse modes. The inverse cloud generator is a large number of cloud droplet distributions $Drop(x, \mu)$ in the known cloud and determines the three numerical eigenvalues Ex, En and He of the normal cloud. The output ensemble problem of the classifier completes the quantitative to qualitative conversion. Therefore, this study focuses only on the inverse cloud generator, and the fuzzy density value of the classifier in different sample regions is used as a cloud droplet and input into the inverse cloud generator to generate a cloud that characterises the reliability of the classifier. The three digital eigenvalues (Ex_k, En_k, He_k) of classifier $e_k(k = 1, 2, \cdots, K)$ are calculated as follows:

(1) The mean of G^k is given by

$$\mu^k = \frac{1}{M} \sum_{m=1}^{M} g_m^k. \tag{8}$$

(2) The variance of G^k is given by

$$S_k^2 = \frac{1}{M-1} \sum_{m=1}^{M} \left(g_m^k - \mu^k\right)^2. \tag{9}$$

(3) The expected value of classifier e_k is given by

$$Ex_k = \mu^k. \tag{10}$$

(4) The entropy of classifier e_k is given by

$$En_k = \sqrt{\frac{\pi}{2}} \times \frac{1}{M} \sum_{m=1}^{M} |g_m^k - \mu^k|. \tag{11}$$

(5) The hyper entropy of classifier e_k is given by

$$He_k = \sqrt{S_k^2 - En_k^2}. \tag{12}$$

Fuzzy density reflects the confidence of the classifier for each category and can effectively describe the recognition performance of the classifier to different sample regions. The cloud model for describing the reliability of the classifier is used because the three digital features (Ex, En, He), which are generated by the cloud model using the fuzzy density of the classifier, can reflect not only the average performance of the classifier but also the stability and randomness of the change of the classifier with the changes in the number of sample. Ex represents the average recognition performance of the classifier, and large expected value means good performance of the classifier. En represents the degree of discretisation of the classifier's capability to identify the sample and reflects the stability of the classifier with the

change in the number of sample. Small value of entropy means better stability (i.e. the performance of the classifier). *He* represents the degree of deviation (i.e. randomness) from the normal operation of the classifier. Therefore, small value of hyper entropy means good performance of the classifier and that the recognition performance of each classifier in different sample regions is converted to the corresponding cloud model.

3.3 Reliability of the classifier based on the cloud model

This study defines reliability from two aspects: subjective and objective. The subjective reliability of the classifier is determined by the first digital feature of the cloud model, that is,

$$\alpha_{sk} = \frac{Ex_k}{\sum\limits_{k=1}^{K} Ex_k}. \tag{13}$$

Good average recognition performance of the classifier means high subjective reliability. The objective reliability is determined by the sum of the entropy and hyper entropy of the cloud model. The entropy characterises the fuzziness of the classifier, and the hyper entropy characterises the randomness of the classifier. Therefore, this study considers using the entropy and hyper entropy of the cloud to construct the optimisation model for calculation.

We let $\varepsilon_k = En_k + He_k$ be the sum of the entropy and hyper entropy of the k-th classifier and construct the optimisation model with the smallest sum of the entropy and hyper entropy of all classifiers. In particular, we set the sum of the fuzziness and randomness of all classifiers to be relatively minimal. Wang [50] considered the three eigenvalues of the normal cloud corresponding to different orders of magnitude. Therefore, the square of the objective reliability is used in the current study to constrain the entropy and hyper entropy in constructing the optimisation model of the entropy and hyper entropy based on the cloud model shown in Eq. (14).

$$\begin{cases} (\alpha_{ok}|k\in1,2,\cdots,K) = \text{argmin} \sum\limits_{k=1}^{K} (\alpha_{ok})^2 \varepsilon_k \\ \sum\limits_{k=1}^{K} \alpha_{ok} = 1, \alpha_{ok}\geq0 \\ \quad k = 1,2,\cdots,K \end{cases} \tag{14}$$

We solve Eq. (14) and construct the Lagrangian function as Eq. (15).

$$L(\alpha_{ok},\lambda) = \sum\limits_{k=1}^{K} (\alpha_{ok})^2 \varepsilon_k + \lambda\left(\sum\limits_{k=1}^{K} \alpha_{ok}-1\right) \tag{15}$$

According to the existence conditions of extremum, we obtain

$$\begin{cases} \dfrac{\partial L(\alpha_{ok},\lambda)}{\partial \alpha_{ok}} = 2\alpha_{ok}\varepsilon_k + \lambda = 0 \\ \dfrac{\partial L(\alpha_{ok},\lambda)}{\partial \lambda} = \sum\limits_{k=1}^{K} \alpha_{ok}-1 = 0 \end{cases}, \tag{16}$$

where $k = 1, 2, \cdots K$. Combining the two equations in Eq. (16), we obtain

$$
\begin{cases}
\lambda = \dfrac{-2}{\sum\limits_{k=1}^{K} (\varepsilon_k)^{-1}} \\
\alpha_{ok} = \dfrac{(\varepsilon_k)^{-1}}{\sum\limits_{k=1}^{K} (\varepsilon_k)^{-1}}
\end{cases}
\tag{17}
$$

The reliability α_k^* of classifier e_k is synthetically determined by subjective and objective reliabilities as follows:

$$
\alpha_k^* = \alpha_{sk} + \alpha_{ok}, k = 1, 2, \cdots K.
\tag{18}
$$

Each classifier can use the three eigenvalues obtained from the cloud model to obtain reliability. For K different base classifiers, we obtain $\zeta = \{\alpha_1^*, \alpha_2^*, \cdots, \alpha_K^*\}$, and the value in ζ is normalised by Eq. (19).

$$
\alpha_k = \frac{1}{2}\left(1 + \frac{\alpha_k^* - \alpha^{\min}}{\alpha^{\max} - \alpha^{\min}}\right), k = 1, 2, \cdots, K
\tag{19}
$$

In Eq. (19), $\alpha^{\min} = \min\limits_{1 \le k' \le K} \{\alpha_{k'}^*\}$ and $\alpha^{\max} = \max\limits_{1 \le k' \le K} \{\alpha_{k'}^*\}$. The normalised reliability $\alpha_k \in [0.5, 1]$ of the classifier e_k is obtained. The reliability of the classifier is proportional to the expected value and is inversely proportional to the entropy and hyper entropy of the classifier. Therefore, the requirement that the classifier should characterise the recognition of different regions is satisfied. The reliability of the classifier can be used as the static weight of the classifier to realise the multi-classifier static weighting ensemble. Considering the characteristics of the sample itself, we aim to identify the real-time recognition performance of the classifier with respect to the currently tested sample and reduce the interference of the unreliable decision information to the classification. For this purpose, the decision credibility of the classifier is defined by measuring the separability amongst the posterior probabilities of the classifier.

3.4 Decision credibility of classifier

Under normal circumstances, the classifier can easily identify the category of the query sample when the posterior probability value obtained by the classifier is highly concentrated in a certain class. At this time, the possibility of misjudgement is low and the credibility of the classifier is high. When the distribution of the posterior probability is uniform, the classifier can exhibit difficulty in discriminating the class of the query sample. Furthermore, the possibility of misjudgement is high and the credibility of the classifier is low. In this study, we use the posterior probability distribution to define the decision credibility of the classifier in measuring the real-

time importance of the classifier in the fusion process. The decision credibility β_k of classifier e_k is solved as follows:

(1) The set of posterior probability values obtained by entering query sample x into classifier e_k is denoted as $P_k = \{p_{k1}, p_{k2}, \cdots, p_{kM}\}$.

(2) The components of the posterior probability value P_k are normalised to [0, 1] and then we obtain $P_k^* = \{p_{k1}^*, p_{k2}^*, \cdots, p_{kM}^*\}$. We let p_{kj}^* be the largest posterior probability value in P_k^* and then define that separability τ_k is the sum between the difference of the other posterior probability values in P_k^* and p_{kj}^*. Specifically, the difference is calculated first and then the sum is obtained using Eq. (20).

$$\tau_k = \frac{1}{M-1} \sum_{m=1}^{M} \left(p_{kj}^* - p_{km}^*\right) = \frac{1}{M-1}\left(Mp_{kj}^* - 1\right) \tag{20}$$

According to Eq. (20), the separability τ_k of the classifier is only related to p_{kj}^*, and high proportion of p_{kj}^* in the output component of classifier e_k means large value of τ_k and separability $\tau_k \in [0, 1]$. When $\tau_k = 1$, the maximum decision output value obtained by classifier e_k is 1 and the other output components are 0. At this point, classifier e_k maximises the probability of sample x being correctly identified and the decision credibility reaches a high value. When $\tau_k = 0$, the posterior probability values of classifier e_k are equal. At this time, the probability of misjudgement is high and the decision credibility of classifier e_k is low.

(3) To emphasise the contribution of large τ_k to the decision making and suppress the interference of small τ_k, a function is used to map τ_k to the decision credibility. When the value of τ_k is small, the assignment of decision credibility is as low as possible. As the value of τ_k increases, the decision credibility of the reliable decision output to the fusion decision rapidly enlarges. The mapping value approaches 1 when the value of τ_k increases to a certain range. To meet the requirements, this study uses the curve fitting in the numerical analysis to construct the sigmoid function in describing the change relationship between τ_k and decision credibility β_k as shown in Eq. (21). Each parameter in Eq. (21) is the optimal value obtained after several simulation experiments.

$$\beta_k = \frac{1}{2(1 + \exp(-(14\tau_k - 6)))} + 0.5 \tag{21}$$

In Eq. (21), $\beta_k \in (0.5, 1)$ and $k = 1, 2, \cdots, K$. Figure 2 shows that, when the proportion of the maximum posteriori probability of the classifier is small, the value of τ_k is also small as reflected by the curve; the mapping value β_k is also low. As separability increases, the decision credibility β_k of the classifier e_k also gradually increases. When $\tau_k > 0.3$, the growth degree of decision credibility β_k increases; when the proportion of the maximum posteriori probability of the classifier is large (i.e. when $\tau_k > 0.65$), the credibility of the classifier is extremely high. In addition, the value of decision credibility β_k is stable and approaches 1.

Fig. 2 Relationship between τ_k and β_k
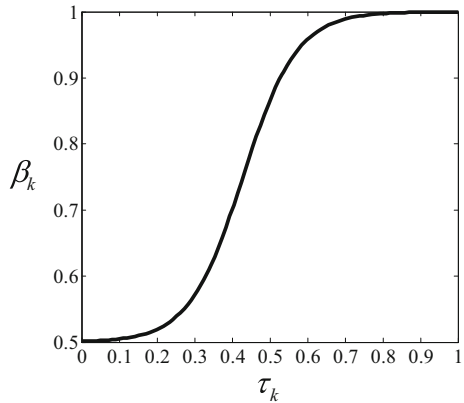
3.5 Description of the proposed algorithm

In this study, a novel method of multi-classifier ensemble based on dynamic weights is proposed. The acquisition of weights is divided into four parts: (1) using different learning algorithms to train data sets to generate base classifiers; (2) using the cross-validation technique to estimate the recognition capability of the classifier in different sample regions, thereby obtaining the three eigenvalues of the cloud model and calculating the reliability of the classifier; (3) using function mapping to obtain the decision credibility of the classifier in accordance with the separability of the output components of the base classifier; (4) the corresponding dynamic weight is assigned to each classifier in accordance with reliability and decision credibility. The obtained weight vector and the output of each base classifier are used to make an integrated judgement to the tested samples. The specific processes are described below:

> Input: data set T_r comprising the samples with known class information, T_r includes M different categories, in which each category contains n samples, a collection of learning algorithms $L = \{l_1, l_2, \cdots l_K\})$ and the sample x to be tested.
> Output: the class of sample x to be tested.

(1) Using the learning algorithm $l_k(k = 1, 2, \cdots, K)$ to train data set T_r, we obtain the base classifier $e_k(k = 1, 2, \cdots, K)$; the stratified cross-validation technique is used to estimate the classification performance of the base classifier e_k, and the obtained confusion matrix is denoted as $CM^k_{M \times M}$.

(2) By calculating the confusion matrix, we obtain the fuzzy density $g^k_m(1 \leq m \leq M)$ of base classifier e_k in different sample regions.

(3) We input g^k_m as a cloud droplet to the inverse cloud generator, obtain the three eigenvalues (Ex_k, En_k, He_k) of the cloud model and calculate the reliability α_k of base classifier e_k.

(4) According to the posterior probability values of sample x to be tested belonging to each class obtained by base classifier e_k, we calculate the decision credibility β_k of the classifier using Eqs. (20) and (21).

(5) We calculate the weight of classifier e_k using Eq. (22) as follows:

$$\omega_k = \frac{\alpha_k + \beta_k}{2}, 0.5 \leq \omega_k \leq 1. \tag{22}$$

(6) We integrate the output components of each base classifier as follows:

$$P_m(x) = \sum_{k=1}^{K} \omega_k p_{km}(x), m = 1, 2, \cdots, M. \tag{23}$$

(7) The result of the integrated discrimination of sample x to be identified is

$$c = \arg \max_{m} \{P_m(x) | m = 1, 2, \cdots, M\}. \tag{24}$$

3.6 Analysis of computational complexity

Computational complexity focuses on the resources required to calculate the problem, such as time and spatial resources. The spatial resources required by the algorithm can meet the requirements of the general computer program. The time complexity of the base classifier for the training process is denoted as $O(Classifier training)$. We calculate the confusion matrix of each base classifier and denote the time complexity of fuzzy density as $O(MKn^2)$, where M represents the number of pattern categories, K represents the number of base classifiers and n represents the number of each training samples. The time complexity of decision recognition is denoted as $O(MK)$. Other processes are ignored. The total time complexity is denoted as $O(O(ClassifierLearn) + n^2MK + MK)$.

4 Application of the proposed algorithm in face recognition

4.1 Feature description of face image

Face recognition [6, 13, 24, 28, 41, 51] uses computers to verify the identity of a face. Face recognition is used to promote many related disciplines and related fields, such as emotion recognition, image process [45] and face detection [3, 22]. Face recognition consists of two steps: feature description and target classification. Commonly used image features include colour, shape and texture features. Amongst them, the texture features based on statistics are favoured by domestic and foreign scholars because of their simplicity and effectiveness; the representative methods are local binary pattern [61], central symmetric local binary pattern [18], local directional pattern [21], local gradient patterns [23], modified census transform [15] and local mean pattern [46]. In this study, the symmetric local graph structure (SLGS) [1] algorithm is used to extract the feature of face images. SLGS is an improvement of the local graph structure [2] operator and is also a local texture description algorithm based on statistics. Compared with the traditional binary pattern, SLGS is no longer limited to the circular neighbourhood of the central pixel, has a range of comparison that can be extended to the comparison between the neighbourhood points and obtains a richer texture feature. However,

when describing the face texture feature, the SLGS algorithm only focuses on the texture information in the horizontal neighbourhood direction whilst ignoring the information in other neighbourhood directions. Accordingly, texture information in other directions is usually lost. The description of the facial features is insufficiently comprehensive. Therefore, SLGS is extended from the directional perspective by adding diagonal and vertical directions and defining 45° SLGS, 90° SLGS and 135° SLGS; the original SLGS referred to herein is 0° SLGS. Figure 3 shows the calculation of the SLGS coding in different directions. Figure 4 shows the SLGS reconstruction diagrams in different directions, including illumination, expression, gestures and occlusion changes. The reconstruction diagram clearly indicates the difference amongst the four different directions of SLGS in the texture feature description.

4.2 Face databases

The stability and validity of the proposed algorithm is tested by conducting simulated experiments on publicly available face databases including ORL, Yale A, FERET and AR databases. These databases cover various conditions of partial occlusion, complex illumination and different expressions and poses. The ORL database contains different images of 40 people; 10 face images are stored for each person, thereby comprising a total of 400 images, including poses, facial expressions and facial ornaments (e.g. glasses) and other changes. Each image is a 112×92 greyscale image. Some of the images in the ORL database are shown in Fig. 5a.

The Yale A database published by Yale University contains 11 images each from 15 individuals with lighting, occlusion and facial emotion variations. Some of the images in the Yale A database are shown in Fig. 5b.

The FERET database contains different images of 200 people, including frontal face images; face images with expression changes; face images under different illumination conditions; face images with left turn and right turn at 15°; face images with left turn and right turn at 30°. Some of the images in the FERET database are shown in Fig. 5c.
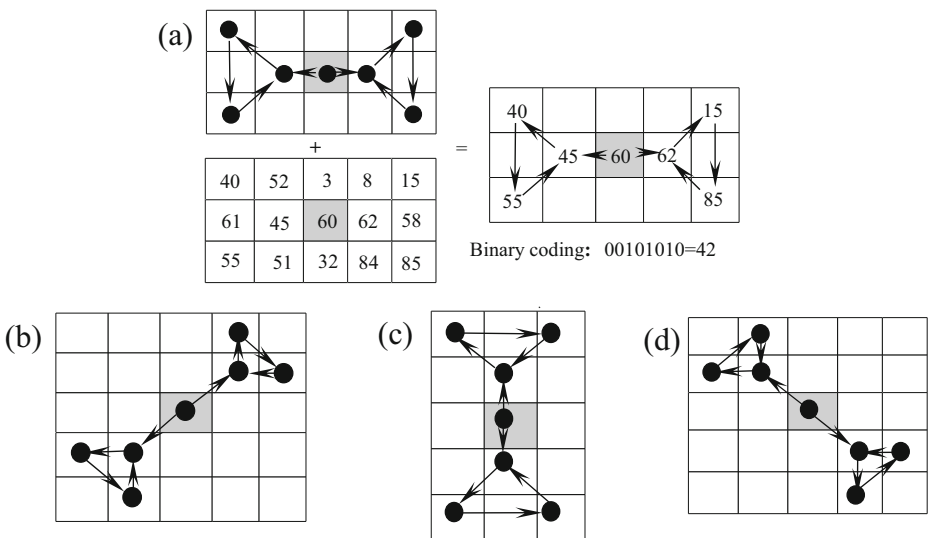

Fig. 3 Calculation of the SLGS coding in different directions: (**a**) 0° SLGS; (**b**) 45° SLGS; (**c**) 90° SLGS; (**d**) 135° SLGS
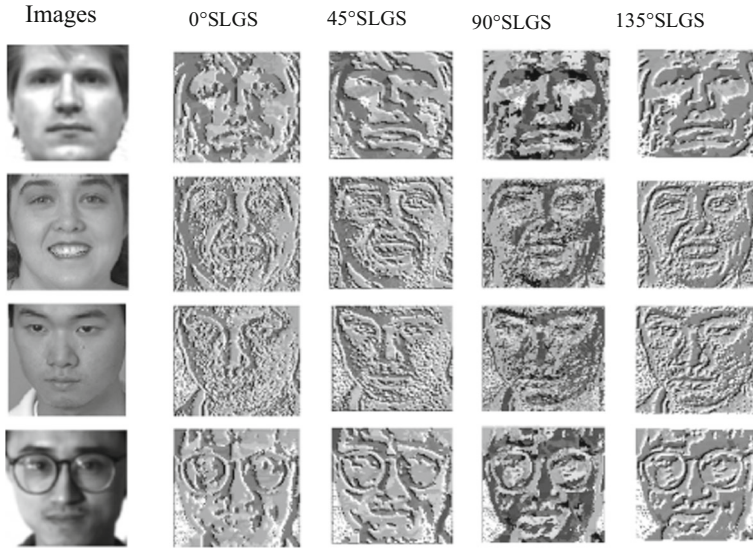
Fig. 4 SLGS reconstruction diagrams in different directions

The AR database consists of more than 4000 frontal face images of 126 persons (70 men and 56 women). Each person has up to 26 images taken in two sessions. The first session contains 13 images, including different facial expressions, illumination variations and occlusions. The second session duplicates the first session two weeks later. To verify the robustness of the method to illumination, occlusion and expression changes, this study divides the AR
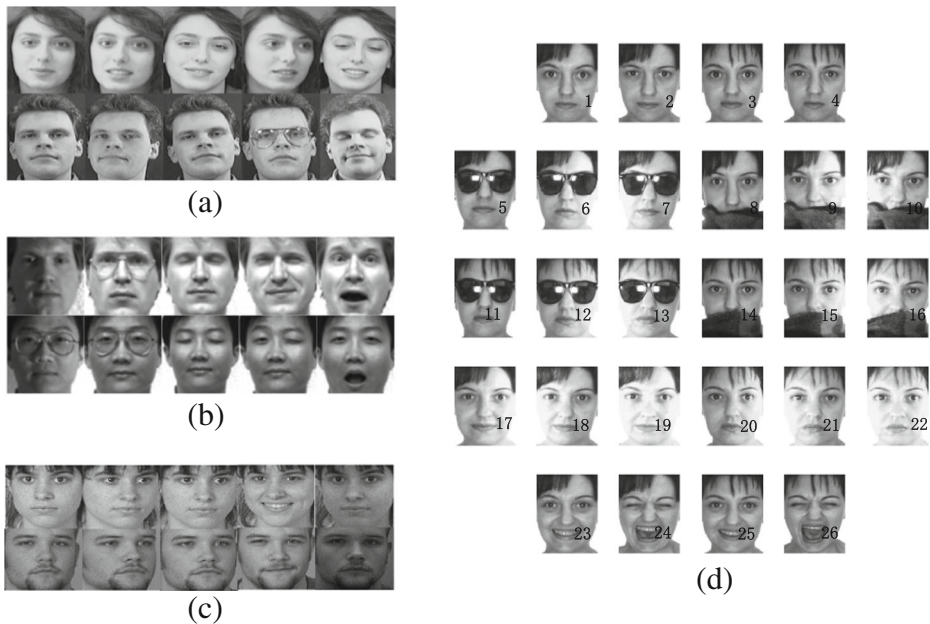

Fig. 5 Face databases: (a) images from the ORL database; (b) images from the Yale A database; (c) images from the FERET database; (d) images from the AR database

database into four parts on the basis of different conditions that contain normal, occlusion, illumination and expression. As shown in Fig. 5d, the images numbered 1–4 are normal face images (the expression is neutral and without illumination changes and occlusion limits), those numbered 5–16 are face images under occlusion conditions and those numbered 23–26 are face images with expression changes. In the experiment, the images in the four face databases are pre-processed to 96 × 96 resolution.

5 Experiments

In the multi-classifier system, this study constructs the base classifier by separately using the nearest neighbour rule and BP neural network combined with the SLGS features in different directions. For the nearest neighbour classifier, the Euclidean distance between the query sample and the training samples in different classes is calculated firstly and then the posterior probability of the nearest neighbour classifier is constructed using the distance. BP neural network is composed of the input, hidden and output layers. In this study, the output of the BP classifier is in the form of posterior probability. The indicator of the performance of the classifier includes recognition accuracy, Kappa coefficient and efficiency. In this study, static ensemble is the fusion result obtained using reliability as the weight of the classifier. Dynamic ensemble refers to the fusion result obtained using reliability and decision credibility as the weight of the classifier. All experiments and comparisons are conducted under the same evaluation protocol.

5.1 Recognition rate comparison of each classifier under different blocks

In the experiment, five face images are selected randomly from ten face images of each individual and used as training set on the ORL database. In the Yale A database, five training samples of each person are randomly selected and used for training and the other images for testing. In the FERET database, four training samples of each person are randomly selected and used for training and the other images for testing. In the AR database, under occlusion conditions, we randomly select six pieces of numbers 5–16 for training and the other as testing samples. Under illumination conditions, three pieces of numbers 17–22 are randomly selected for training and the other as testing samples. Under the condition of expression changes, we randomly select two pieces of numbers 23–26 as facial expression image for training and the other as testing samples. Figure 6 shows the recognition results of the classifiers trained by four different algorithms and the proposed ensemble methods under different blocks.

Overall, in different face databases, the recognition performance firstly shows a rising trend and then a flat or descending trend with the increase in the number of blocks. Compared with the single classifier, the proposed multi-classifier fusion methods present good performance advantages regardless of their use in static and dynamic ensemble. Their recognition rates are also much higher than that of single classifier prior to fusion. Among them, the classification effect of the dynamic ensemble is better than that of the static ensemble. For the feature description algorithm, large number of blocks in a certain range means good details of the spatial texture structure described and good performance of the classifier. Given the increase in the number of blocks, the number of the pixels in different sub blocks gradually reduces and the operator's feature description capability gradually weakens. Consequently, the classifier's identification performance firstly rises and then declines. Figure 6 also shows that the classification performance of the classifier formed by different characteristics of the same training
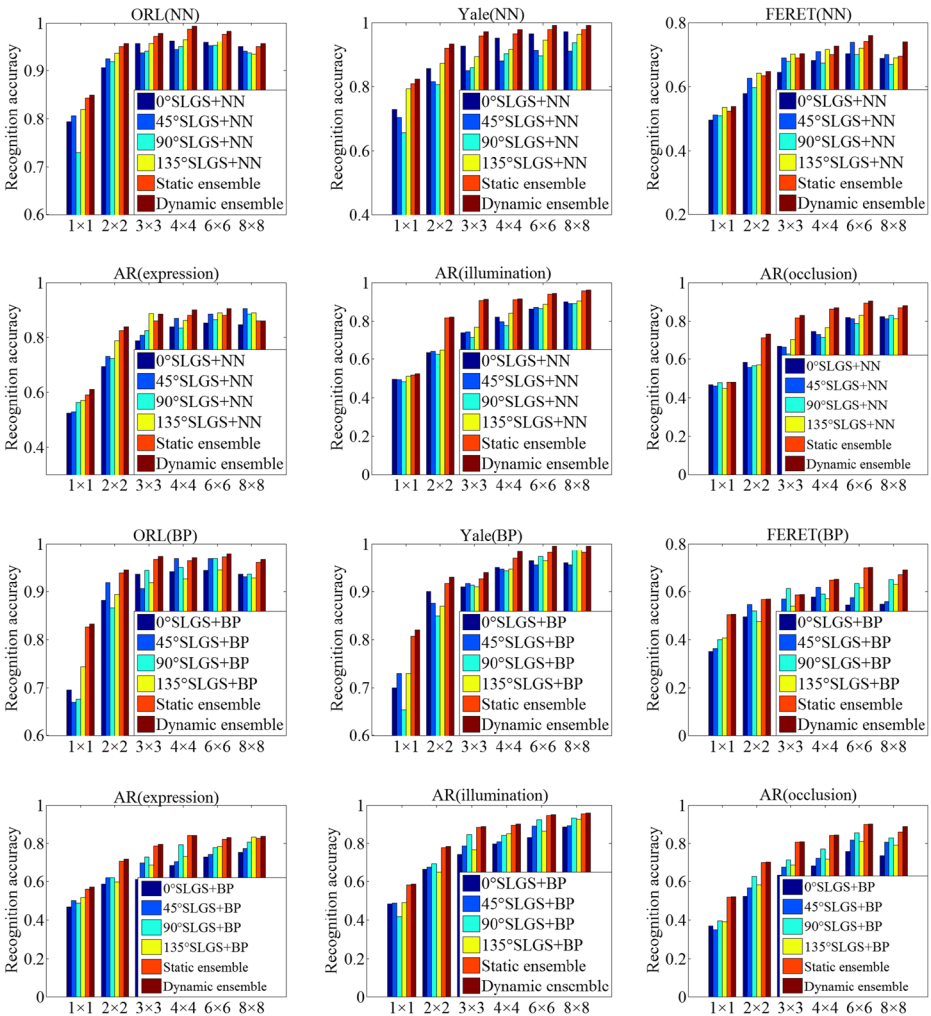
Fig. 6 Recognition rates of the classifiers trained by four algorithms and the proposed ensemble methods under different blocks

sample differs. By taking the FERET database as an example, the recognition performance of 0° SLGS + NN is in a state of equality with 90° SLGS + NN on the FERET database, whereas the recognition performance of 0° SLGS + NN is inferior to that of the 45° SLGS + NN. Such a scenario is also reflected from the side that the recognition capability of different classifiers to identify the same sample area differs. This study complements the missing information of original SLGS from the perspective of direction, and the classifiers trained by four different texture description operators construct the integrated system with the difference. In addition, the importance of the classifier in the fusion decision is described on the basis of the prior knowledge and posteriori output of each classifier. The three eigenvalues of the cloud model can fully describe the reliability of the classifier in different sample regions and can fully utilise the classification advantages of the classifier. The decision credibility can effectively describe the real-time performance of the

classifier's recognition performance with the change in the number of sample to be tested. Thus, the final dynamic ensemble achieves good recognition performance that is higher than that of static ensemble and single classifier. The increase in the number of blocks can also increase time complexity. Notably, this study aims to find a compromise between "simplicity" and "effectiveness." In Fig. 6, when the number of blocks is 6×6, the system obtains the best recognition performance with a small time cost. Therefore, this parameter is used in the subsequent experiments to address the corresponding face image.

Kappa coefficient to further evaluate the optimal classification accuracy of dynamic and static ensembles. Kappa coefficient represents the proportion of the error reduction produced by the evaluated classification to that of the completely random classification. Large Kappa coefficient value means good performance of the classifier. In the contrast experiment, we select 13 images of each person as training samples on the AR database and the remaining samples as testing samples. The number of training samples on other face databases remains unchanged. Table 1 shows the comparison of the Kappa coefficients of static and dynamic ensemble methods.

Table 1 shows that the Kappa coefficients of the dynamic weighted ensemble method are higher than those of the static method on different databases. It can be seen from the definition of Kappa coefficient that the higher the Kappa value, the higher the accuracy of classification. Obviously, the method of dynamic weighted ensemble is better than the static method. In the process of weight acquisition, compared with the static method, the dynamic method not only considers the recognition ability of the classifier in different sample areas, but also takes the difference of the samples into consideration and considers the recognition result of the current target by this classifier as an important indicator in measuring the importance of it. Thereby, the acquired weight can effectively reduce the interference of unreliable decision information to fusion and thus obtaining improved classification performance.

5.2 Recognition rate comparison of each classifier under different number of training samples

In further verifying the effectiveness of the proposed algorithm, 2–8 images per person in the ORL and Yale A databases are used for training and the remaining samples for testing. In the FERET database, 2–5 training samples are selected randomly and the other images are for testing. In the AR database, 9–14 training samples are randomly selected and the other images are for testing. Following the conclusion of experiment (Section 5.1), the images in four face databases are blocked into 6×6 blocks. We repeat the experiments four times and calculate the average recognition rate for each database. The experimental results are shown in Fig. 7.

Figure 7 shows the curves of the recognition rates of the base classifiers based on four algorithms and the proposed multi-classifier fusion methods. These curves change with the

Table 1 Comparison of the Kappa coefficients (%) of static and dynamic ensembles

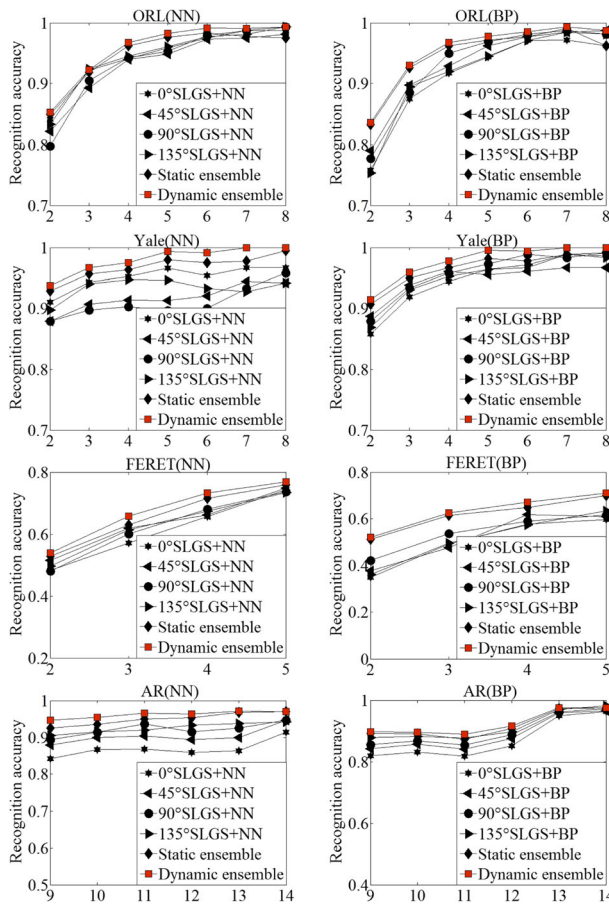| Kappa	ORL	Yale A	FERET	AR
Static ensemble	98.08	98.57	72.36	96.97
Dynamic ensemble	98.97	1	73.37	97.12

Fig. 7 Curves of the recognition rates of the base classifiers and the proposed methods under different training samples

increase in the number of training samples. In a certain range, large number of training samples results in complete classifier training and high recognition rate. In different face databases, the multi-classifier system composed of four different texture features and using NN or BP significantly improves after the decision fusion. Therefore, the method is valid. The reasons are that the four different texture features are portrayed from different directions in the feature description and that the classifier that comprises them is highly complementary. In the decision-making fusion, the dynamic ensemble exhibits a good description of the classifier's reliability in different sample areas and real-time performance with the change in number of sample compared with the static ensemble. The former method also can be more targeted to play the recognition advantage of each classifier in different regions and different samples and ultimately improve the overall recognition rate. In the FERET and AR databases, the 0° SLGS + NN and 0°SLGS + BP classifiers present a lower performance than do those in the vertical and diagonal directions. By contrast, the performance of the four classifiers is comparable on the ORL and Yale A databases. The experimental results show that different directions of SLGS play different roles in describing the image texture features. Thus, SLGS should be extended from a directional perspective.

5.3 Recognition rate and identification time comparison of different multi-classifier ensemble methods

In this experiment, the optimal recognition rate obtained by the proposed dynamic ensemble method is compared with the recognition rate obtained by several typical fusion methods (e.g. voting method, maximum rule, D-S fusion rule and fuzzy integral method). The number of training samples of the ORL and Yale A databases is 5, that of the FERET database is 4, that of the AR database is 13 and the remaining samples are used as testing samples. The comparison results are shown in Table 2.

The experimental data in Table 2 indicate that the recognition performance of the maximum rule method, D-S fusion method, fuzzy integral method and dynamic ensemble method is better than that of the voting method. The voting method is a hard-grading standard, which considers that the input pattern either belongs to a certain class, or not. The voting method is an abstract-level fusion method, while the other fusion algorithms are belonging to metric-level fusion which contains more decision information.Thus they can achieve higher recognition rates than voting method. In the ORL and Yale A databases, the recognition performance of the maximum rule method, D-S fusion method and fuzzy integral method is close to that of the proposed method. On the contrary, the difference is significant in the AR database. Such a difference mainly occurs because the changes in illumination, occlusion and expression are complex in the AR database, thereby possibly causing different degrees of interference and reducing the performance of the fusion decision. D-S evidence theory is based on the assumption that different pieces of evidence should remain independent, thereby requiring that the classifiers should be independent of one another. However, in practice, complete independence is difficult. Fuzzy integral avoids the limitation of D-S evidence theory that the evidence needs to be completely independent, thus, it can effectively deal with the uncertainty of the classification decision. However, fuzzy integral generally makes the training performance index of the classifier as the basis for determining the fuzzy density, which no longer change once determined. When the classifier outputs unreliable information to the decision-making layer, decision-making interference occurs. Thus, the integrated classification effect becomes undesirable. Multi-classifier weighted ensemble can effectively emphasise the decision-making contribution made by excellent classifiers and suppress the influence of unreliable information output by high-deviation classifiers on classification. Thus it can reduce the interference of unreliable decision factors and still maintain its fusion decision-making performance above or close to that of the optimal classifier in the multi-classifier system.

Table 2 Comparison of optimal recognition rates (%) under different integration rules, mean processing time per image (milliseconds for MATLAB implementation in an i7 quad core processor with 4 GB RAM, windows 7 32bit)

Methods		Voting	Maximum rule	D-S fusion rule	Fuzzy integral	Proposed method
ORL	Acc	94.84	96.85	97.81	97.46	98.28
	T	182.4	64.9	66.4	88.6	68.9
Yale A	Acc	91.33	96.67	97.78	98.67	99.34
	T	127.9	55.9	61.7	84.1	56.7
FERET	Acc	69.94	71.5	72.75	71.25	73.34
	T	274.1	75.4	80.4	94.5	80
AR	Acc	92.57	93.78	96.92	95.38	97.17
	T	562.5	80.0	89.6	108.3	86.4

Therefore, the proposed method exhibits a stronger resistance to unreliable decision-making interference than common fusion methods do in the fusion decision making.

With regard to process time, that of different algorithms from fast to slow is in the following order: maximum rule method, dynamic ensemble method, D-S fusion method, fuzzy integral method and the voting method. The voting method consumes much time to find the nearest sample, especially when the training set is large. The fuzzy integral involves an iterative process and thus consumes much time when dealing with multiple types or large sample problems. The D-S fusion method presents increased time complexity during the construction of its basic probability distribution function, which reduces the classification efficiency. Given that the weights of the dynamic ensemble method are assigned by defining reliability and credibility, its reliability can be obtained by offline training and its credibility can be obtained by dispersion mapping of the classifier with negligible the time consumption for mapping. As a result, the dynamic integration method exhibits good efficiency in classification, which is only slightly below that of the maximum rule method. Considering the classification accuracy and efficiency of the proposed method in face recognition, its feasibility and effectiveness are verified.

To better illustrate the advantages of our method, the optimal experimental results of the multi-classifier dynamic ensemble method in face recognition are compared with those of some existing face recognition methods.

As shown in Table 3, the proposed multi-classifier dynamic ensemble method achieves a good classification performance. The classifier adjusts its weight by analysing its recognition capability in different regions of the sample set and the decision information of the output, thereby emphasising the "contribution" of the excellent classifier to the final classification result. Accordingly, some error-prone divisions can be corrected and the interference of unreliable decision output to fusion can be decreased.

6 Discussion

In a multi-classifier system, the base classifier should satisfy the difference and complementarity. If all the classifiers incorrectly identify the sample, then the ensemble will be meaningless. Therefore, a classifier set with difference and complementarity should be comprehensively constructed. In classifying different query samples, the performance of the classifiers shows a certain difference. If a numerical value is used to measure its importance,

Table 3 Recognition rates (%) compared with other methods

Methods	ORL	Yale A	FERET	AR
SRC [51]	92.30	–	33.07	88.44
PNN [24]	94.95	–	56.87	95.82
TWSBF + LDA [20]	97.15	–	59.45	97.51
LGP + BAW-GSA [6]	90.7	96.5	–	–
W-LBPAT [61]	96.5	94.91	–	–
FMSD [28]	97.56	96.53	53.46	–
ACS-IDA [41]	95	97.8	86.4	–
Proposed method	98.28	99.34	73.34	97.17

then the value should reflect not only the identification performance of the base classifier in the entire sample space but also the adaptability of the base classifier, which changes with the change in the number of query sample. This study aims to dynamically adjust the weights of each base classifier. Consequently, the weight of the classifier providing the correct classification information increases and the weight of the classifier providing incorrect information decreases. The weight exhibits strong adaptability and can be updated as the query sample changes.

7 Conclusions

In this study, we analyse the advantages and disadvantages of existing weighted ensemble methods on the basis of the study on ensemble learning. A dynamic weight assignment method that defines the reliability and credibility of the classifier is proposed. Compared with general methods, two aspects, namely, the performance on the training set and the decision information output when identifying specific target, are considered in acquiring the weights of the classifier. This method can thus describe the importance of the classifier comprehensively. The comparison between the results of simulation experiment on face recognition and those of the existing popular ensemble methods proves that the proposed method can effectively reduce the interference of unreliable decision information and efficiently improve the recognition performance of the classification system. The current study on multi-classifier ensemble method is of practical value to face recognition problem, of methodological significance to other fields in addition to machine learning and possesses a wide application prospect.

Issues related to multi-classifier ensemble methods, such as constructing and selecting efficient base classifiers and obtaining effective results of the base classifier, should be explored. In this study,the proposed method is only verified on standard face databases,corresonding optimisation and adjustment should be made according to the specific circustances when solving other pattern classification problems.Meanwhile, the base classifiers constructed in this paper are homogeneous, this would inevitably lead to the overlapping of wrong samples, which will affect the classification system to make accurate decision.The next research focus is how to screen the classifier by measuring the difference between the classifiers so as to obtain a better classification result.

References

1. Abdullah MFA, Sayeed MS, Muthu KS et al (2014) Face recognition with symmetric local graph structure (SLGS). Expert Syst Appl 41(14):6131–6137
2. Abusham EEA, Bashir HK (2011) Face recognition using local graph structure (LGS). Human-computer interaction. Interaction techniques and environments. Springer Berlin Heidelberg, pp 169–175

3. Ban Y, Kim SK, Kim S et al (2014) Face detection based on skin color likelihood. Pattern Recogn 47(4): 1573–1585
4. Bejani M, Gharavian D, Charkari NM (2014) Audiovisual emotion recognition using ANOVA feature selection method and multi-classifier neural networks. Neural Comput & Applic 24(2):399–412
5. Bhimani J, Mi N, Leeser M et al (2017) FiM: performance prediction model for parallel computation in iterative data processing applications. IEEE International Conference on Cloud Computing
6. Chakraborti T, Chatterjee A (2014) A novel binary adaptive weight GSA based feature selection for face recognition using local gradient patterns, modified census transform, and local binary patterns. Eng Appl Artif Intell 33(1):80–90
7. Chen B, Wang S, Jiao LC (2015) PolSAR image classification method based on weighted majority vote ensemble. Huazhong Univ of Sci & Tech(Natural Science Edition) 43(3):79–82
8. Cheon Y, Kim D (2009) Natural facial expression recognition using differential-AAM and manifold learning. Pattern Recogn 42(7):1340–1350
9. Cho SB, Kim JH (1995) Combining multiple neural network by fuzzy integral for robust classification. IEEE Trans Syst Man Cybern 25(2):380–384
10. Codella N, Cai J, Abedini M et al (2015) Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. Machine learning in medical imaging, pp 118–126
11. Cui J, Liu Y, Xu Y et al (2013) Tracking generic human motion via fusion of low- and high-dimensional approaches. IEEE Trans Syst Man Cybern Syst 43(4):996–1002
12. Dai J, Liu H, Wang M et al (2017) Research and application of an uncertainty concept model: Time-varying cloud model. IEEE, International Conference on Cloud Computing and Big Data Analysis
13. Ding C, Choi J, Tao D et al (2016) Multi-directional multi-level dual-cross patterns for robust face recognition. IEEE Trans Pattern Anal Mach Intell 38(3):518–531
14. Eleftheriadis S, Rudovic O, Pantic M (2014) Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition. IEEE Trans Image Process 24(1):189–204
15. Froba B, Ernst A (2004) Face detection with the modified census transform. Proceedings of 6th International Conference on Automatic Face and Gesture Recognition. Seoul, South Korea: IEEE 91–96
16. Gao H, Yang Z, Bhimani J et al (2017) AutoPath: harnessing parallel execution paths for efficient resource allocation in multi-stage big data frameworks. International Conference on Computer Communications and Networks
17. Guo K, Li W (2011) Combination rule of D-S evidence theory based on the strategy of cross merging between evidences. Expert Syst Appl 38(10):13360–13366
18. Heikkilä M, Pietikäinen M, Schmid C (2009) Description of interest regions with local binary patterns. Pattern Recogn 42(3):425–436
19. Hu ZP, Li J, Zh SH (2013) Sub-modular sparse representation algorithm for robust pattern recognition based on Borda voted weighting. Chin J Sci Instrum 34(10):2309–2315
20. Huang ZH, Li WJ, Wang J et al (2015) Face recognition based on pixel-level and feature-level fusion of the top-level's wavelet sub-bands. Inf Fusion 22:95–104
21. Jabid T, Kabir MH, Chae O (2010) Local directional pattern (LDP) —a robust image descriptor for object recognition. Proceedings of 7th International Conference on Advanced Video and Signal Based Surveillance. Boston, MA: IEEE, pp 482–487
22. Jin J, Xu B, Wang Y (2015) A face detection and location method based on feature binding. Signal Process Image Commun 36:179–189
23. Jun B, Kim D (2012) Robust face detection using local gradient patterns and evidence accumulation. Pattern Recogn 45(9):3304–3316
24. Kumar R, Banerjee A, Vemuri BC et al (2011) Maximizing all margins: pushing face recognition with kernel plurality. International Conference on Computer Vision. IEEE Computer Society, pp 2375–2382
25. Kuncheva LI, Rodríguez JJ (2014) A weighted voting framework for classifiers ensembles. Knowl Inf Syst 38(2):259–275
26. Kwak KC, Pedrycz W (2005) Face recognition: a study in information fusion using fuzzy integral. Pattern Recogn Lett 26(6):719–733
27. Lee C-C, Mower E, Busso C et al (2011) Emotion recognition using a hierarchical binary decision tree approach. Speech Comm 53(9):1162–1171
28. Li X, Song A (2013) Letters: fuzzy MSD based feature extraction method for face recognition. Neurocomputing 122:266–271
29. Li DY, Meng HJ, Shi XM (1995) Membership clouds and membership cloud generators. Computer R & D 32(6):15–20
30. Li G, Zhang Z, Wang L et al (2017) One-class collaborative filtering based on rating prediction and ranking prediction. Knowl-Based Syst 124:46–54

31. Liu HJ, Liu Z, Jiang WL et al (2010) Approach based on cloud model and vector neural network for emitter identification. Acta Electron Sin 38(12):2797–2804
32. Liu Y, Cui J, Zhao H et al (2012) Fusion of low-and high-dimensional approaches by trackers sampling for generic human motion tracking. International Conference on Pattern Recognition. IEEE, pp 898–901
33. Liu Y, Nie L, Han L et al (2015) Action2Activity: recognizing complex activities from sensor data. International Conference on Artificial Intelligence. AAAI Press, pp 1617–1623
34. Liu Y, Nie L, Liu L et al (2016) From action to activity: sensor-based activity recognition. Neurocomputing 181:108–115
35. Liu L, Cheng L, Liu Y et al (2016) Recognizing complex activities by a probabilistic interval-based model. Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, pp 1266–1272
36. Liu Y, Zhang L, Nie LQ et al (2016) Fortune teller: predicting your career path. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAI-16, pp 201–207
37. Liu Y, Liang Y, Liu S et al (2016) Predicting urban water quality with ubiquitous data. arXiv:161009462
38. Liu Y, Zheng Y, Liang YX et al. Urban water quality prediction based on multi-task multi-view learning. Proceedings of the 25th International Joint Conference on Artificial Intelligence
39. Lu Y, Wei Y, Liu L et al (2017) Towards unsupervised physical activity recognition using smartphone accelerometers. Multimed Tools Appl 76(8):10701–10719
40. Luo Y, Chen S, He X et al (2013) Alphanumeric character recognition based on BP neural network classification and combined features. Int J Comput Intell Syst 6(6):1108–1115
41. Naik MK, Panda R (2016) A novel adaptive cuckoo search algorithm for intrinsic discriminant analysis based face recognition. Appl Soft Comput 38(C):661–675
42. Orrite C, Rodríguez M, Martínez F et al (2008) Classifier ensemble generation for the majority vote rule. Progress in Pattern Recognition, Image Analysis and Applications 5197:340–347
43. Owusu E, Zhan Y, Mao QR (2014) A neural-AdaBoost based facial expression recognition system. Expert Syst Appl 41(7):3383–3390
44. Preoţiuc-Pietro D, Liu Y, Hopkins D et al (2017) Beyond binary labels: political ideology prediction of twitter users. Meeting of the Association for Computational Linguistics, pp 729–740
45. Ren F, Li B, Chen Q (2013) Single parameter logarithmic image processing for edge detection. Ieice Trans Inform Syst 96(11):2437–2449
46. Ren FJ, Li YQ, Xu LF et al (2016) Face recognition method based on local mean pattern description and double weighted decision fusion for classification. J Image Graph 21(5):565–573
47. Santana MC, Marsico MD, Nappi M et al (2016) MEG: texture operators for multi-expert gender classification. Comput Vis Image Underst 156:4–18
48. Song YJ, Li DY, Yang XZ et al (2000) Reliability evaluation of electronic products based on cloud models. Acta Electron Sin 28(12):74–76 68
49. Sun B, He J (2015) Weighted joint sparse representation-based classification method for robust alignment-free face recognition. J Electron Imaging 24(1):013018
50. Wang XF, Xiao MS (2010) Approach of group decision making based on normal distribution interval number with incomplete information. Control and Decision 25(10):1494–1498
51. Wright J, Yang AY, Ganesh A et al (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210
52. Xu Q (2013) A novel machine learning strategy based on two-dimensional numerical models in financial engineering. Math Probl Eng (2):1–6
53. Xu L, Krzyzak A, Suen CY (1992) Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Trans Cybern 22(3):418–435
54. Xu Q, Wu J, Chen Q (2014) A novel mobile personalized recommended method based on money flow model for stock exchange. Math Probl Eng 2014. https://doi.org/10.1155/2014/353910
55. Xu Q, Wang Z, Wang F et al (2017) Thermal comfort research on human CT data modeling. Multimed Tools Appl. https://doi.org/10.1007/s11042-017-4537-9
56. Yang Z, Awasthi M, Ghosh M et al (2017) A fresh perspective on total cost of ownership models for flash storage in datacenters. IEEE International Conference on Cloud Computing Technology and Science
57. Yang Z, Tai J, Bhimani J et al (2017) GReM: dynamic SSD resource allocation in virtualized storage systems with heterogeneous IO workloads. PERFORMANCE Computing and Communications Conference. IEEE
58. Yang Z, Wang J, Evans D et al (2017) AutoReplica: automatic data replica manager in distributed caching and data processing systems. PERFORMANCE Computing and Communications Conference. IEEE
59. Yang J, Li J, Liu S (2017) A novel technique applied to the economic investigation of recommender system. Multimed Tools Appl (8):1–16

60. Zhang YS, Guo J (2012) Word sence disambiguation based on ensemble classifier with dynamic weight adaptation. J Chin Inform Process 26(1):3–9
61. Zhang JY, Zhao HP, Chen S (2014) Face recognition based on weighted local binary pattern with adaptive threshold. J Electron Inform Technol 36(6):1327–1333

Fuji Ren He received the Ph.D. degree in 1991 from Faculty of Engineering, Hokkaido University, Japan. He worked at CSK, Japan, where he was a chief researcher of NLP from 1991. From 1994 to 2000, he was an associate professor in Faculty of Information Science, Hiroshima City University. He became a professor in Faculty of Engineering, the University of Tokushima in 2001. His research interests include Artificial Intelligence, Language Understanding and Communication, and Affective Computing. He is a member of the IEICE, CAAI, IEEJ, IPSJ, JSAI, AAMT, a senior member of IEEE, and a Fellow of The Japan Federation of Engineering Societies.

Yanqiu Li She was born in Huaibei, China, in 1988. She is a Ph.D. student of Hefei University of Technology. She received her Bachelor degree from Chaohu University in 2010, and her Master degree from Hefei University of Technology in 2014. Her research interests include Humanoid Robots, Affective Computing, and Computer Vision.

Min Hu She received her Ph.D. degree in 2004 from Hefei University of Technology, China. From 1994 to 1999, she did some research in the Institute of Microcomputer, and from 2000 to now, she has been doing some teaching and research work in School of Computer and Information in Hefei University of Technology. She became a professor in School of Computer and Information in 2008. Her research interests include Video image processing and Affective Computing.