Misclassified Samples based Hierarchical Cascaded Classifier for Video Face Recognition

Zheyi Fan, Shuqin Weng, Yajun Zeng, Jiao Jiang, Fengqian Pang, Zhiwen Liu
School of Information and Electronics, Beijing Institute of Technology
Beijing, China
[e-mail: funye@bit.edu.cn]
*Corresponding author: Zheyi Fan

Received July 2, 2016; revised September 30, 2016; accepted October 25, 2016; published February 28, 2017

Abstract

Due to various factors such as postures, facial expressions and illuminations, face recognition by videos often suffer from poor recognition accuracy and generalization ability, since the within-class scatter might even be higher than the between-class one. Herein we address this problem by proposing a hierarchical cascaded classifier for video face recognition, which is a multi-layer algorithm and accounts for the misclassified samples plus their similar samples. Specifically, it can be decomposed into single classifier construction and multi-layer classifier design stages. In single classifier construction stage, classifier is created by clustering and the number of classes is computed by analyzing distance tree. In multi-layer classifier design stage, the next layer is created for the misclassified samples and similar ones, then cascaded to a hierarchical classifier. The experiments on the database collected by ourselves show that the recognition accuracy of the proposed classifier outperforms the compared recognition algorithms, such as neural network and sparse representation.

Keywords: Video face recognition, scatter, misclassified samples, clustering, hierarchical cascaded classifier

https://doi.org/10.3837/tiis.2017.02.009 **ISSN : 1976-7277**

1. Introduction

Face recognition is a fundamental research topic in the field of computer vision, artificial intelligence and pattern recognition. It has been widely used in many applications such as intelligent surveillance, human-computer interaction and public safety. Video face recognition utilizes the motion information and numerous face features. It can not only prevent disguised invading but also enhance the recognition accuracy of individuals, thus has attracted considerable attention [1-5]. At present, video face recognition can be categorized into classifier-based, spatial-temporal model-based and match-based algorithms. The classifier-based recognition algorithms mainly use classifiers like sparse representation [6,7], support vector machine (SVM) [8] and hidden Markov model (HMM) [9], etc. Chen et al. [6] optimize the process of creating dictionary, which adds the motion information into the dictionary, making full use of the various features to achieve the robust face recognition. Wolf et al. [8] propose the SVM-minus algorithm to calculate the similarity between the two video sequences. Kim et al. [9] learn the dynamic procedure by HMM, which considers the spatial-temporal feature in order to improve the recognition accuracy.

A common problem that is encountered by all the methods mentioned above, is that their face recognition algorithms rely on the single classifier, which is sensitive to the data distribution. However, due to the variation of posture, expression and illumination existed in video, the face data of the same person can be quite different while those for different persons can be similar, which means the practical face data in videos are widely spread. Thus, the face recognition with single classifier is inclined to cause the problem that the within-class scatter is higher than the between-class scatter, resulting in the unsatisfactory recognition accuracy and a poor generalization ability.

Therefore, many researchers apply the cascaded classifier to the detection and recognition tasks [10-11]. Liang et al. [10] use a cascade of boosting classifier to select a small number of features from a huge feature set, and efficiently apply it in vehicle detection. The cascaded classifier is also used in video face recognition [12-15]. Connolly et al. [13] introduce the incremental learning based on dynamic particle swarm optimization to the process of cascading classifiers, which can cascade the classifiers by selecting partial optimal particles. Hassanpour et al. [14] propose an Ensemble of Abstract Sequence Representatives (EASR) method in sequences to reduce the influence of noise and redundant information in video face recognition. Yang [15] applies the cascading to the spare representation-based classifier, reducing the computational complexity in face recognition. There are also some latest works [16-18] focus on finding better representations of images and exploring the local structure of images, which can be applied to face classification. Li et al. [16] integrate image understanding and feature learning into a joint learning framework to reduce the semantic gap, which can deal with different image understanding tasks like clustering and classification. Tang et al. [17] propose a

discriminant hashing function by exploiting local discriminative information such as local scatter to implement approximate similarity search. However, the classical cascaded classifiers like Ada-Boost [19] and Bagging [20], ignore the samples' subspace distribution, making the selection of samples random and blind, which leads to the instability of the classifiers.

Aiming at the problems mentioned above, this paper proposes a hierarchical cascaded classifier for video face recognition. Motivated by the unsatisfactory face recognition accuracy of the existed algorithms, this paper focuses on dealing with the problems that the within-class scatter is higher than the between-class scatter and the randomness in the selection of samples. So as to enhance the recognition accuracy, the proposed algorithm focuses on the subdivision of the misclassified samples and their similar samples to maximally avoid misclassification. Firstly, the video sequence of each person is divided into several clusters, and the first-layer classifier is created by computing the minimum of the mean distance. Secondly, the rest samples in the gallery are recognized by the single layer classifier. If the recognition result is wrong, the next-layer classifier is created for the misclassified samples and their similar samples. At last, the final classifier is completed by cascading hierarchical classifiers. The hierarchical cascaded classifier then can be applied to recognizing face in the query.

The main contributions of this paper can be summarized as follows: (i) As a cascaded classifier, the proposed classifier solves the problem widely existed in face recognition that the within-class scatter is higher than between-class scatter by clustering. (ii) By creating multi-layer classifier for misclassified samples and their similar samples, the classifier figures out the problem of randomly selecting samples. (iii) This paper applies a novel face recognition algorithm which is capable for robust human face recognition via the proposed hierarchical cascaded classifier.

The rest of this paper is organized as follows. In Section 2, the detailed algorithm of creating a single classifier is provided. Section 3 introduces the misclassified samples based hierarchical cascaded classifier. The video face recognition algorithm via the proposed classifier is described in Section 4 and is conducted on the practical databases in Section 5. Finally, our discussion concludes this paper in Section 6.

2. Single classifier created by clustering

As face image in video has the differences in posture and illumination, the within-class scatter is higher than between-class scatter. If the original data is to be handled without preprocessing, the recognition results will be unsatisfactory. To enhance the total recognition accuracy, this paper designs the first layer of the classifier, which can be regarded as a single classifier, to minimize the within-class scatter by clustering. The feature extraction and clustering of the single classifier are discussed in details in the following parts.

2.1 Feature extraction and analysis

The face features are of high dimension and information redundancy. Principal Component Analysis (PCA) [21] is a widely used tool for dimension reduction. Therefore, this paper utilizes PCA to extract the face features. The training samples are firstly converted to gray images and then we extract the gray scale characteristics from them. The face data distribution obtained by PCA is shown in **Fig. 1**. To make a concise illustration, **Fig. 1** only shows 6 kinds of face and the former 2 dimensions of feature, distinguishing different categories of the samples by colors and shapes.

In order to explain the problem in video face recognition, we introduce two scatter matrices. The within-class scatter matrix measures the scatter of samples in each class around the mean value of that class. The between-class scatter matrix, on the other hand, measures the scatter of class-conditional expected values around the global mean. Within-class scatter S_w and between-class scatter S_b are defined as follows:

$$S_w = \sum_{c=1}^{N} \sum_{i=1}^{n_c} \left(x_i^c - m^c \right) \left(x_i^c - m^c \right)^{\mathrm{T}} \tag{1}$$

$$S_b = \sum_{c=1}^{N} n_c \left(m^c - m \right) \left(m^c - m \right)^{\mathrm{T}} \tag{2}$$

where N is the number of classes, n_c is the total number of the c-th kind samples, m^c is the mean of all the c-th kind samples. m is the mean of all samples and x_i^c is the i-th sample of the c-th kind.

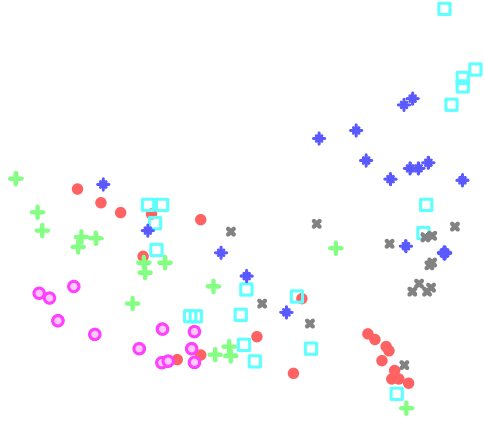


Fig. 1. The data distribution of 6 kinds face samples

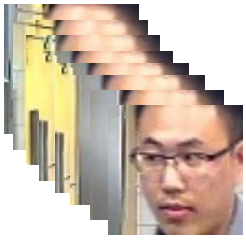
Within-class scatter is the variance of the samples belonging to the same kind, reflecting the degree of distribution concentration of the same kind. The bigger the within-class scatter is, the more disperse the data distribution is. Between-class scatter is

the variance of all kinds of samples, reflecting the distribution similarity of the data belonging to the different kinds. The smaller the between-class scatter is, the more similar the data distributes and the nearer the distribution center is. As shown in **Fig. 1**, the two kinds of samples indicated by dark blue and light blue are widely spread, while their distribution regions are similar. The data belonging to the same kind is disperse due to the differences in posture and illumination, and the data belonging to the different kinds distributes similarly under the similar posture and illumination. As a result, there is a very serious problem in face recognition that the within-class scatter is higher than between-class scatter and thus the recognition accuracy is influenced.

2.2 Clustering and single-classifier creating

To enhance the total recognition accuracy, the single classifier is created to minimize the within-class scatter by clustering. Clustering analysis is an unsupervised data analysis method. It clusters the similar samples according to the distance of samples. This paper utilizes hierarchical clustering to cluster samples. It computes the number of classes by analyzing distance tree firstly and then clusters the similar samples.

The samples in gallery are classified into different classes according to the features extracted by PCA, and we define the feature matrix and label as $\{x^i, y^i\}(i=1,2,\ldots,N)$, where N is the total number of classes. There are k persons in the training data, and the data belonging to the same person is defined as $\{X_t\}(t=1,2,\ldots,k)$. The process of clustering is shown in **Fig. 2**. The left is a video sequence and the right shows the distance tree of all samples, where the samples with same color belong to one cluster. And the clustering result is in **Fig. 3**.



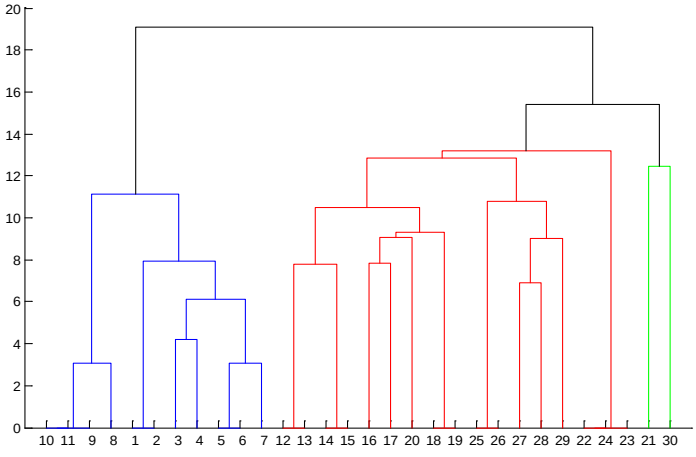


Fig. 2. The distance tree of samples

(a) The first cluster

(b) The second cluster

(c) The third cluster

Fig. 3. The clustering results of one person

The dataset after clustering is defined as $\left\{S_t^j=\left\{x_{t,j}^m\right\}\right\}\left(j=1,2,\cdots,n_t\right)\left(m=1,2,\cdots,\left|S_t^j\right|\right)$. It represents the samples set of the j-th cluster of the t-th people, where m is the number of samples in this cluster and $n_t\left(t=1,2,\cdots,k\right)$ is the number of subclasses created by clustering. Classes are labeled with $A,B,\cdots,K,\cdots$, and the framework of the first layer is shown in **Fig. 4**.

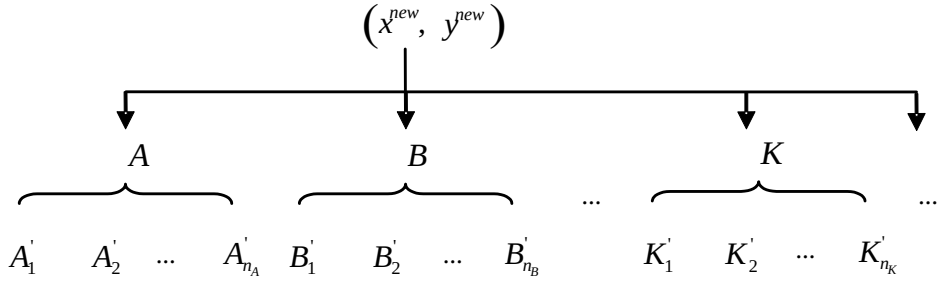


Fig. 4. The first-layer framework

In the training process, we firstly choose one video sequence of each person for clustering and create the single-layer classifier. Then, we recognize the other video sequences in gallery by computing the minimum distance, defining those testing samples as $\left\{x^{new},y^{new}\right\}$. The concrete procedure is detailed as **Algorithm 1**.

Algorithm 1 Minimum Distance Method

Input: The testing samples $\{x^{new}, y^{new}\}$, the clustering center $\{S_t^j = \{x_{t,j}^k\}\}(j = 1,2,\ldots,n_t)$ $(k = 1,2,\ldots,\left|S_t^j\right|)$, distance vector $dis = \sum_{t=1}^{k} n_t \times 1$.

for $t = 1,2,\ldots,k$
  sum = 0;
  for $j = 1,2,\ldots,n_t$
    count = 0;
    for $k = 1,2,\ldots,\left|S_t^j\right|$
      count++;
      sum = sum+$\left|x_{t,j}^k - x^{new}\right|$;
    end
    sum/count $\rightarrow dis$;
  end
end

Output: $j^* = \arg\min_{j=1,2,\ldots n_t} dis$, t^* representing the class of j^*.The class of the testing sample is defined as the class of the clustering center with the minimum distance from it.

The data distribution concentrates after clustering and the value of within-class scatter is effectively decreased. The single-layer classifier constructed by minimum distance method can deal with the condition of all subclasses while has the low computational cost, making it suitable for video face recognition.

3. Cascaded classifier designed for video face recognition

A cascaded classifier merges several precise single classifiers for recognition. Because of its better generalization ability, the cascaded classifier has been widely used in pattern recognition. Nevertheless, the traditional cascaded classifier has not taken the distribution of samples into consideration, selecting samples randomly and blindly. Inspired by the principle of learning the misclassified samples in Ada-Boost, this paper creates a hierarchical cascaded classifier based on misclassified samples and their similar samples to select samples. The procedure of training the hierarchical cascaded classifier is detailed as follows:

We recognize the rest face samples in the gallery using **Algorithm 1** and then analyze the result. If $t^* = y^{new}$, namely the recognition result is right, we add this sample to the cluster $S_{t^*}^{j^*} = S_{t^*}^{j^*} \bigcup x^{new}$. Otherwise, we create a next-layer classifier for this cluster. The construction of the second-layer classifier is shown in **Fig. 5**. The new sample is denoted as

$\left(x^{new}, y^{new}\right)$, being recognized according to the minimum of distances and the result is the n_K-th cluster of the K-th class. In this case, the recognition result K is different from the sample label y^{new}, namely the recognition result is wrong. Besides, the sample x^{new} is similar with the samples belonging to the n_K-th cluster of K. So we create the next-layer classifier for the misclassified sample x^{new} and the cluster K'_{n_K} that is similar with the x^{new}.

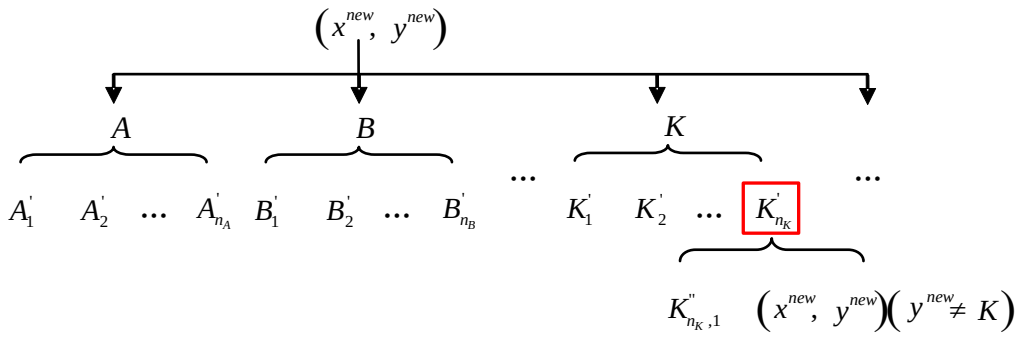


Fig. 5. The construction of the next-layer classifier for misclassified samples

We recognize all samples in the gallery and create the next-layer classifier for the misclassified samples and their similar samples by the procedure above. The final hierarchical cascaded classifier is shown in **Fig. 6**. In the second layer of the classifier, the class with only one label, for example, $A''_{2,1}$, is the same class as the original cluster A'_2 in the upper layer. The other classes have two labels, for example, $A_2D''_2$, in which the former part represents the recognition result A'_2 in the upper layer and the latter part represents the true class D'_2 respectively. The composite label $A_2D''_2$ means that the samples belonging to this class is misclassified in the upper layer as A'_2, and these samples actually belong to the class D'_2, which is already existed in the upper layer. The labels in the under layers are named regularly as the second layer. The samples classified into the same branch distribute similarly, thus they may have the similar expression, posture, illumination, etc. After the hierarchical classifier completed layer by layer, the SVM is applied in the last layer to achieve the multi-class classification.

In the recognition process, the hierarchical cascaded classifier is utilized to recognize the face data in the query. The cluster is determined by minimum distance method, and the final output of class is obtained by SVM.

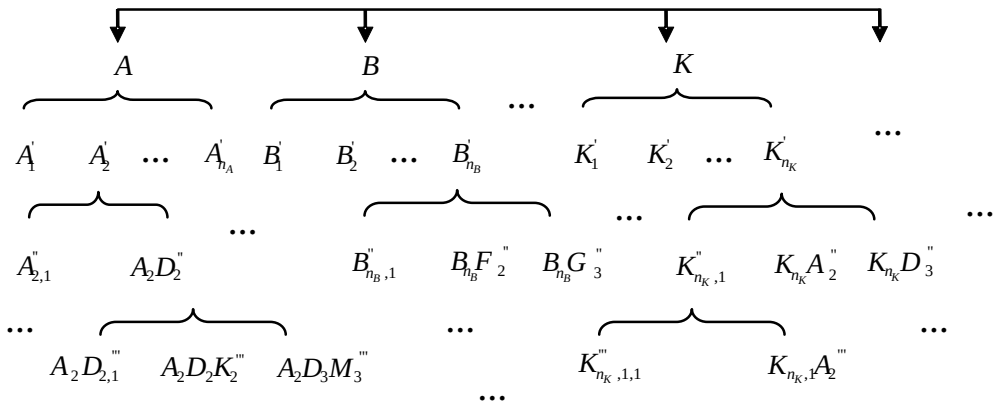


Fig. 6. The framework of hierarchical cascaded classifier

Take the binary classification for example, the training process and the recognition process are elaborated in **Fig. 7** and **Fig. 8** respectively. All the video sequences of the two classes are divided into two parts and the face feature is extracted by PCA. One part of the samples is used to cluster and create the first-layer classifier. The data distribution of the two classes is shown in **Fig. 7(a)**, where the same color stands for the same class and the different shapes stand for different subclasses. The other part of the samples is used to recognize and create the next-layer classifier. The new training samples belonging to the red class distribute as the red triangle in **Fig. 7(b)**. The result of the new samples recognized by the first-layer classifier is shown in **Fig. 7(c)**, in which the samples distributing close to the two subclasses of red samples are recognized as the red class, while those close to the blue star samples are recognized as the blue class. The purple line shows the classification hyperplane of the first layer. Obviously, the blue triangle samples are recognized wrong till now. Thus, the next-layer classifier is to create for the misclassified samples and their similar ones. As to distinguish the misclassified samples with the red samples in the first layer, the misclassified samples in **Fig. 7(d)** are recolored with green. The orange line shows the classification hyperplane of the second layer. The training process of the classifier is finished, and the output of this process is the trained classifier, shown as the classification hyperplane.

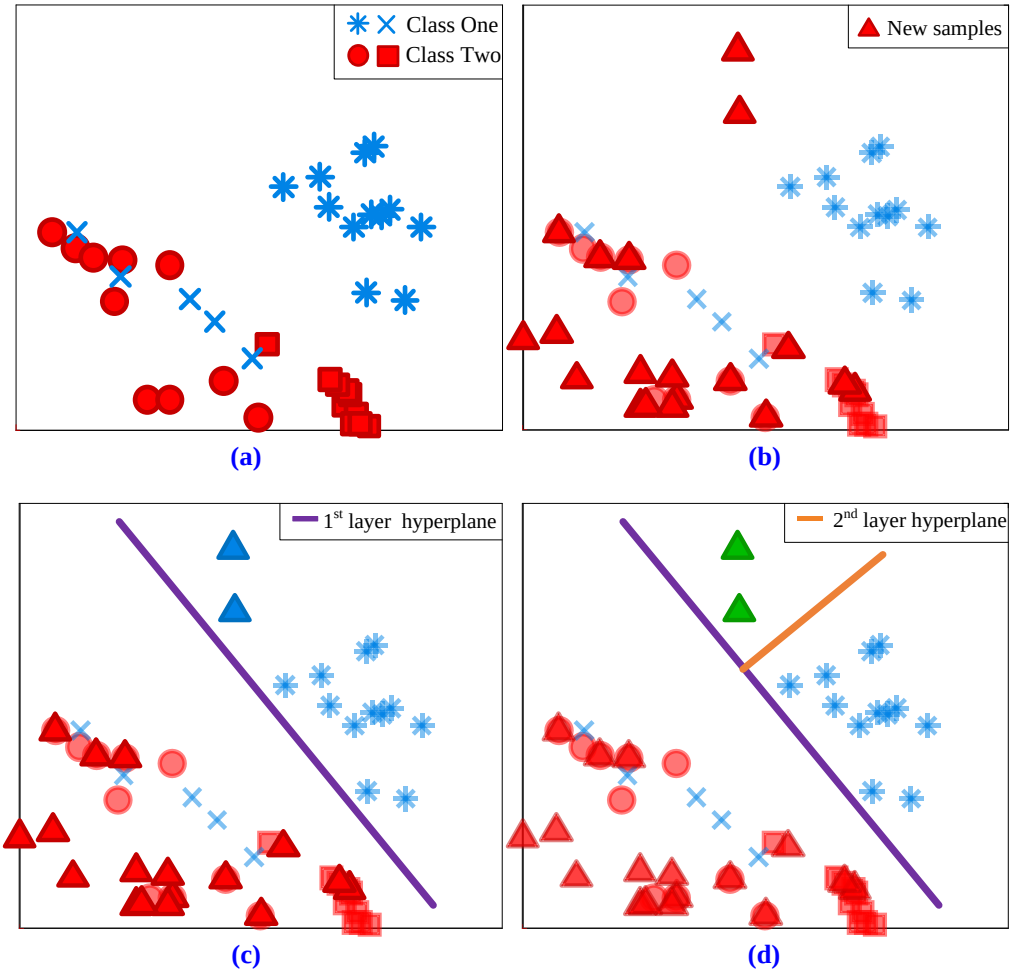


Fig. 7. The training process of binary classification

The recognition process is illustrated by **Fig. 8**. The testing samples belonging to the red class distribute as the black diamond in **Fig. 8(a)**. The recognition result obtained by the trained 2-layer classifier is shown in **Fig. 8(b)**. The samples classified as red are recognized right directly by the first layer, while others are recognized by the two layers of the classifier and the final recognition result is green class. Because the green class is the subclass for the misclassified samples, whose real class is the same as the red, the recognition result is right.

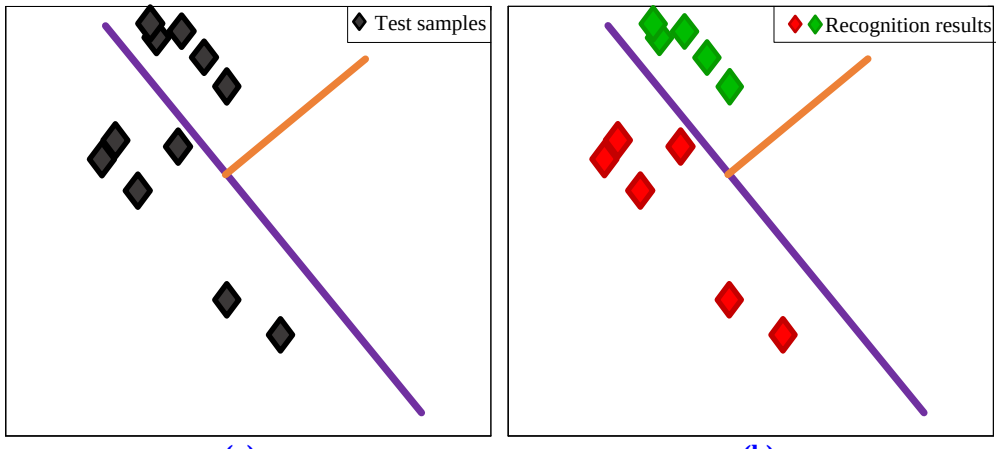


Fig. 8. The recognition process of binary classification

The single classifier created by clustering in **Section 2** concentrates the data distribution. It effectively reduces the within-class scatter with low computational cost, which contributes to the calculating speed of the whole algorithm. However, it just computes the number of classes by analyzing distance tree and clusters the similar samples roughly. Thus, the hierarchical clustering is utilized in the next-layer classifier construction in this section.

The classifier created by hierarchical cascading in this way can have different training processes for different kinds of samples. The samples which are similar to the clusters in the first layer are easy to be recognized correctly so the number of training layers is less. However, if the samples are difficult to be decided in the first layer, the next-layer classifier will be created for them. Moreover, in the process of creating next-layer classifier, there is no influence on the previous classifiers and the data distribution in the former layers. The only impact is on the misclassified samples and their similar samples, that is creating the next-layer classifier for them to achieve an accurate recognition result. In a word, the more layers created for the similar samples, the higher the recognition accuracy is.

In the recognition process, the procedure and the time consumption of different samples, like the easily-misclassified samples and the easily-recognized samples, are significantly different. For the easily-recognized samples, the number of training layers is less, so the recognition time is less. On the contrary, the easily-misclassified samples are trained by more layers, so the recognition time increases. The cascaded classifier has different layers for different data, thus it can reduce the recognition time and maintain the recognition accuracy meanwhile.

This hierarchical cascaded classifier focuses on creating multi-layer classifiers for misclassified samples. By this way, it avoids randomly and blindly choosing samples in traditional classifier, and considerably enhances the generalization ability.

4. Video face recognition algorithm

This paper applies a novel video face recognition algorithm which is capable of robustly recognizing human face via the proposed hierarchical cascaded classifier. There are several main steps in this process, including clustering and computing the minimum distance firstly, and cascading classifiers created in the next layer for the misclassified samples. The holistic procedures are divided into training process and recognition process, and the details of the algorithm are explained in **Algorithm 2**.

Algorithm 2 Video Face Recognition Algorithm

Input:

Gallery: The feature matrix and label of the face video sequences $\left\{x^i, y^i\right\}\left(i=1,2,\ldots,n\right)$

Query: The samples of the test video sequences $\left\{x^{test}, y^{test}\right\}$

Training process:

for each person k do

Select one video sequence in gallery to cluster using the algorithm in **Section 2**

Recognize other sequences $\left\{x^i, y^i\right\}\left(i=1,2,\ldots,n_{left}\right)$ using the **Algorithm 1**

if $t^* \neq y^i$ then

Create the next-layer classifier as **Section 3**

end if

end for

Output: The whole hierarchical cascaded classifier

Recognition process:

for each sample j in query do

Recognize j by the hierarchical cascaded classifier as elaborated in **Section 3**

end for

Output: The label of the testing sample, namely the recognition result

5. Experiments and analysis

In order to explain the efficiency of our method, we evaluate the proposed algorithm on both our own database and some of the public databases, compared with the baseline algorithms and the new methods respectively.

5.1 Experiments on the practical data and analysis

The experiments in this section are conducted on the practical data collected by ourselves. The videos have recorded 20 persons coming into the door, containing motion information and appearance changes. The gallery and query are both video sequences. For each person, there are 3 sequences as gallery and 1 sequence as query. In order to keep balance of data size, we fix the number of frames in one video sequence to 30 and the part of training data

is listed in **Fig. 9**. Every video sequence in query has no less than 30 frames, however, the number of frames is not fixed, the total of which is 801 frames of the 20 persons.

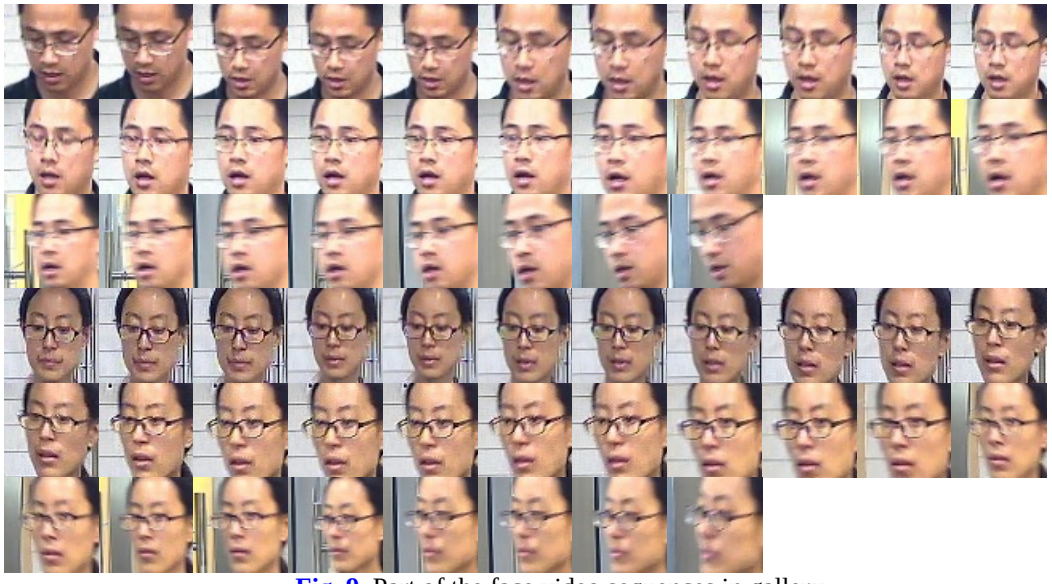


Fig. 9. Part of the face video sequences in gallery

There are two steps in the training process: feature extraction and classifier generation. We extract 40-dimensional face features by PCA. And classifier generation can be divided into two procedures: (i) one video sequence for each person is selected to cluster to generate the first-layer classifier, and (ii) the rest samples in the gallery are recognized and generate the hierarchical classifier layer by layer for the misclassified samples. The two-layer classifier of the first person is shown in **Fig. 10**.

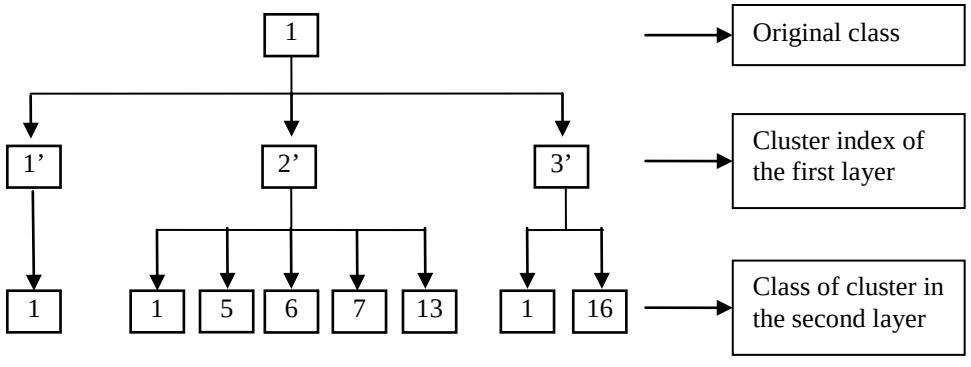


Fig. 10. The two-layer cluster structure of the first person

The computational time, the number of layers and the recognition accuracy are three important factors of the cascaded classifier. In the recognition process, the recognition

accuracy is defined as the number of right recognized frames divide by the total of frames. The relationship of these factors is listed in the **Table 1**.

Table 1. The relationship among the number of layers, recognition accuracy and computational time

Num. of layers	Recognition accuracy	Computational time (s)
1	63.42%	0.09
2	75.28%	0.91
3	83.27%	1.42
4	89.64%	1.76
5	91.51%	1.94

Obviously, the more layers created, the higher recognition accuracy is. As the number of layers is increasing, the architecture of the cascaded classifier is more complete, the learning of the similar samples is more accurate, the recognition accuracy is improved.

At the same time, the computational time for the training is increasing with the growing number of layers. When there is only one layer, the recognition is finished by computing the minimum distances, which indicates the minimum computational time. When there is more than one layer, the multi-layer classifier generation contains both the minimum distances computation and the recognition by SVM, so it costs more time. As analyzed by **Table 1**, the computational time is increasing, however, the total time cost is not that long.

In order to explain the influence on the relationship between the within-class scatter and the between-class scatter achieved by the hierarchical cascaded classifiers, the relationship of the within-class scatter and between-class scatter in computed in **Table 2**. (i) When there are no layers, the data is not clustered to many subclasses, so the data are widely spread and the within-class scatter is even larger than the between-class scatter. (ii) When there is one layer, the data is clustered to subclasses so the data distribution is becoming intensive and the within-class scatter is smaller than the between-class scatter. (iii) When there are more layers, the classifier is cascaded by creating next layers, and with the growing number of layers, the ratio of between-class scatter to within-class scatter is larger. In a word, the recognition result indicates that the clustering and the hierarchical cascaded classifiers can reduce the within-class scatter and raise the between-class scatter.

Table 2. The relationship among the number of layers, within-class scatter and between-class scatter

Layers No.	Within-class scatter	Between-class scatter	Ratio
0	2.57×10^4	1.55×10^4	0.60
1	1.74×10^4	3.03×10^4	1.74
2	2.41×10^4	5.06×10^4	2.10
3	2.71×10^4	5.98×10^4	2.21
4	2.76×10^4	6.63×10^4	2.40

The recognition accuracy of our proposed approach and the sparse representation, neural network, SVM is compared in **Table 3**. The experimental results show that our method outperforms the compared video face recognition algorithms.

Table 3. The recognition accuracy of different algorithms

Algorithm	Recognition accuracy
Sparse representation (vote)	81.15%
Sparse representation (mean)	77.53%
Neural network	73.28%
SVM	89.26%
Our proposed classifier (5 layers)	**91.51%**

5.2 Comparison with the state-of-the-art

In this section, we present the comparison with the state-of-the-art methods. We select two public face databases, Face Recognition Technology (FERET) [22] and Extended Yale B [23] as benchmark datasets. The face images were all resized to 32×32.

The comparison with the similar work [15] is conducted on the Extended Yale B dataset, which consists of 2414 frontal-face images of 38 individuals. The data of each person is captured under various laboratory-controlled lighting conditions. The classifier in [15] is called Cascade Sparse Representation-based Classifier (CSRC), which is the improvement of Sparse Representation-based Classifier (SRC) [24]. Both the CSRC and our misclassified samples based hierarchical cascaded classifier utilize PCA in face feature extraction. For each subject, we randomly select half of the images as gallery for training (i.e., about 32 images per subject), and the other half for testing. We extract the face features and use PCA to reduce the features into different dimensions, which are the same as in work [15], and each experiment is tested about 30 times, obtaining the recognition accuracy by averaging. **Table 4** reports the comparison of the recognition accuracy of the SRC, CSRC and our proposed classifier under the conditions of different feature dimensions.

Table 4. The recognition accuracy of different algorithms under different feature dimensions

Feature dimensions	SRC	CSRC	Our proposed classifier (5 layers)
30	90.90%	**91.60%**	91.30%
84	95.50%	97.30%	**97.50%**
150	96.80%	98.20%	**98.90%**
300	98.30%	99.30%	**99.50%**

The results in the second and fourth columns are compared to show that our proposed classifier outperforms the SRC, while the third and fourth columns show that when face features are extracted in higher dimension, our method also presents better performance than the CSRC. Moreover, when the feature dimension is high enough, our 5-layer classifier shows the almost perfect accuracy. Due to the data in the Extended Yale B has diversity of illumination, our proposed classifier can robustly create the cascaded classifier based on misclassified samples. Thus, it has higher recognition accuracy than the compared methods.

There are also some other state-of-the-art methods in face recognition field, for example, Xuan et al. [25] propose the structure scatter-based multi-scale patch classifier with subclass representation (MSPSCRC), which is built on the collaborative representation-based classification (CRC) [26] and multi-scale patch-based CRC (MSPCRC) [27]. We use the FERET and Extended Yale B face databases to test the proposed method, and the creating of the cascaded classifier is the same as the way in **Section 5.1**. In this comparison, the baselines CRC, MSPCRC, MSPSCRC, patch based CRC (PCRC), subclass representation-based classification (SCRC), k-nearest neighbor (KNN) methods and patch-based nearest neighbor (PNN) classifier, were used for comparison. **Table 5** and **Table 6** present the comparison of our best results and the recognition accuracy of these algorithms [25] on the FERET face database and the Extended Yale B face database respectively.

Table 5. The recognition accuracy of different algorithms on the FERET face database

Algorithm	Recognition accuracy	Algorithm	Recognition accuracy
CRC	66.39%	MSPSCRC	74.44%
SCRC	74.36%	KNN	64.47%
PCRC	29.60%	PNN	58.92%
MSPCRC	31.40%	**Our proposed classifier (5 layers)**	**85.20%**

Table 6. The recognition accuracy of different algorithms on the Extended Yale B face database

Algorithm	Recognition accuracy	Algorithm	Recognition accuracy
CRC	73.76%	MSPSCRC	93.17%
SCRC	82.57%	KNN	64.44%
PCRC	94.00%	PNN	64.18%
MSPCRC	95.49%	**Our proposed classifier (5 layers)**	**96.20%**

Both the results in the **Table 5** and **Table 6** show that the proposed classifier achieves the highest recognition accuracy among all the compared algorithms in the two databases.

6. Conclusion

This paper proposes a novel hierarchical cascaded classifier based on misclassified samples and applies it to video face recognition. It gathers the data by clustering analysis so as to reduce the within-class scatter and raise the between-class scatter. Moreover, by creating multi-layer classifier for misclassified samples and their similar samples, this classifier solves the problem of randomly selecting samples. The experimental results strongly demonstrate that the proposed classifier has significantly better performance than the compared face recognition algorithms.

References

[1] H. S. Bhatt, R. Singh and M. Vatsa, "On rank aggregation for face recognition from videos," in *Proc. of 20th IEEE Int. Conf. on Image Processing*, pp. 2993-2997, Sep. 15-18, 2013. Article (CrossRef Link)

[2] W. Xu, S. Lee and E. Lee, "A Robust Method for Partially Occluded Face Recognition," *KSII Transactions on Internet and Information Systems*, vol. 9, no. 7, pp. 2667-2682, July, 2015. Article (CrossRef Link).

[3] E. G. Ortiz, A. Wright and M. Shah, "Face recognition in movie trailers via mean sequence sparse representation-based classification," in *26th IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3531-3538, June 23-28, 2013. Article (CrossRef Link)

[4] Z. W. Huang, R. P. Wang, S. G. Shan and X. L. Chen, "Projection Metric Learning on Grassmann Manifold with Application to Video based Face Recognition," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 140-149, June 7-12, 2015. Article (CrossRef Link)

[5] R. G. Cinbis, J. Verbeek and C. Schmid, "Unsupervised metric learning for face identification in TV video," in *Proc. of IEEE Int. Conf. on Computer Vision*, pp. 1599-1566, Nov. 6-13, 2011. Article (CrossRef Link)

[6] Y. C. Chen, V. M. Patel, P. J. Phillips and R. Chellappa, "Dictionary–based face recognition from video," in *Proc. of IEEE European Conf. on Computer Vision*, pp. 766-779, Oct. 7-13, 2012. Article (CrossRef Link)

[7] H. Zheng, Q. Ye and Z. Jin, "A Novel Multiple Kernel Sparse Representation based Classification for Face Recognition," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 4, pp. 1463-1480, Apr., 2014. Article (CrossRef Link).

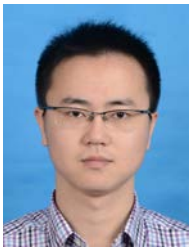
[8] L. Wolf and N. Levy, "The SVM-minus similarity score for video face recognition," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3523-3530, June 23-28, 2013. Article (CrossRef Link)

[9] M. Kim, S. Kumar, V. Pavlovi and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1-8, June 23-28, 2008. Article (CrossRef Link)

[10] P. Liang, G. Teodoro, H. Ling, E. Blasch, G. Chen and L. Bai, "Multiple kernel learning for vehicle detection in wide area motion imagery," in *15th IEEE Int. Conf. on Information Fusion (FUSION)*, pp. 1629-1636, July 9-12, 2012.

[11] H. Wang and Y. Cai, "Monocular based road vehicle detection with feature fusion and cascaded Adaboost algorithm," *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 22, pp. 3329-3334, Nov. 2015. Article (CrossRef Link)

[12] X. O. Tang and Z. F. Li, "Video based face recognition using multiple classifiers," in *Proc. of the 6th IEEE Int. Conf. on Automatic Face and Gesture Recognition*. pp. 345-349, May 17-19, 2004. Article (CrossRef Link)
[13] J. F. Connolly, E. Granger and R. Sabourin, "An adaptive ensemble of fuzzy ARTMAP Neural Networks for video-based face classification," in *IEEE Congress on Evolutionary Computation*, pp. 1-8, July 18-23, 2010. Article (CrossRef Link)
[14] N. Hassanpour and L. Chen, "A hierarchical training and identification method using Gaussian process models for face recognition in videos," in *11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition*, pp. 1-8, May 4-8, 2015. Article (CrossRef Link)
[15] Y. Yang, "Face recognition algorithm based on cascade sparse representation-based classifier," *Industry & Mine Automation*, vol. 40, no. 5, pp. 46-48, May. 2014.
[16] Z. C. Li, J. Liu, J. H. Tang, and H. Q. Lu, "Robust Structured Subspace Learning for Data Representation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 10, pp. 2085-2098, Oct. 2015. Article (CrossRef Link)
[17] J. H. Tang, Z. C. Li, M. Wang and R. Z. Zhao, "Neighborhood discriminant hashing for large-scale image retrieval," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2827-2840, Sep. 2015. Article (CrossRef Link)
[18] Z. C. Li, J. Liu, Y. Yang, X. F. Zhou and H. Q. Lu, "Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection," *IEEE Transactions on Knowledge & Data Engineering*, vol. 26, no. 9, pp. 2138-2150, Sep. 2014. Article (CrossRef Link)
[19] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System and Science*, vol. 55, no. 1, pp. 119-139, Jan. 1997. Article (CrossRef Link)
[20] G. Fumera, F. Roli and A. Serrau, "A theoretical analysis of Bagging as a linear combination of classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1293-1299, July 2008. Article (CrossRef Link)
[21] G. Fumera, F. Roli and A. Serrau, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, Jan. 1991. Article (CrossRef Link)
[22] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000. Article (CrossRef Link)
[23] A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 23, no. 6, pp. 643-660, June 2001. Article (CrossRef Link)
[24] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 31, no. 2, pp. 210-227, Feb. 2009. Article (CrossRef Link)
[25] S. Xuan, S. Xiang and H. Ma, "Subclass representation-based face-recognition algorithm derived from the structure scatter of training samples," *IET Computer Vision*, vol. 10, no. 6, pp. 493-502, Apr. 2016. Article (CrossRef Link)
[26] L. Zhang, M. Yang and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *IEEE International Conference on Computer Vision*, pp. 471-478, Nov. 6-13, 2011. Article (CrossRef Link)
[27] P. Zhu, L. Zhang, Q. Hu and S. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *European Conference on Computer Vision*, pp. 822-835, Oct. 7-13, 2012. Article (CrossRef Link)



Zheyi Fan is a lecturer at the School of Information and Electronics, Beijing Institute of Technology. He received his B.S., M.S. and Ph.D. degrees from Beijing Institute of Technology, Beijing, China, in 2003, 2007 and 2013 respectively all in Electronic Engineering. His current research interests include image processing and pattern recognition, video object tracking, multimodal biometric identity recognition and crowd behavior analysis. He has already authored or coauthored over 20 papers in the above areas.

Shuqin Weng received her B.S. degree in Information and Communication Engineering from Beijing Institute of Technology, Beijing, China, where she is currently pursuing a master's degree. Her current research interests include moving object detection and shadow detection, action recognition and pedestrian detection based on deep learning, which focuses on crowded scenes.

Yajun Zeng received her M.S. degree in Electronic Science and Technology from Beijing Institute of Technology, Beijing, China, in March 2016. Her main research interests focus on multimodal biometric-based identity recognition technologies in video sequences, including video face recognition based on sparse representation and human gait recognition based on discrete cosine transform.

Jiao Jiang received her B.S. degree in Information and Communication Engineering from Beijing Institute of Technology, Beijing, China, where she is currently pursuing a master's degree. Her current research interests include motion pattern learning and crowd behavior analysis based on clustering algorithm, visual object tracking based on particle filter, both of which focus on complex scenes and abrupt motion.

Fengqian Pang received his B.S. degree in Communication Engineering from Civil Aviation University of China, Tianjin, China, in 2011, and his M.S. degree in Information and Communication Engineering from Beijing Institute of Technology, Beijing, China, in 2013. He is currently pursing Ph.D. degree in Beijing Institute of Technology, China. His main research interests include computer vision and medical image analysis.

Zhiwen Liu is a professor at the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. He received his B.S. degree from Xidian University, Xi'an, Shaanxi, China, in 1983, and his M.S. and Ph.D. degrees from Beijing Institute of Technology, China, in 1986 and 1989 respectively all in Electronic Engineering. His research interests include array signal processing, biomedical signal and image processing, smart wearable medical devices, estimation and detection theory. He has already authored or coauthored over 200 papers in the above areas. He is a senior member of the Chinese Institute of Electronics (CIE) and a member of IEEE.