An Ensemble Convolutional Echo State Networks for Facial Expression Recognition*

Guihua Wen, Huihui Li, Danyang Li
School of Computer Science and Engineering
South China University of Technology
Guangzhou, China
crghwen@scut.edu.cn, 29777562@qq.com, 525231151@qq.com

***Abstract*—Facial expressions recognition (FER) plays a much important role in various applications from human-computer interfaces to psychological tests. However, most methods are confronted with the quality of the face images, vanishing gradients problem, over-trained problem, difference of face images such as in age and ethnicity, mulitple parameters required tuning, and dubious class labels in the training data. These negative factors largely hurt the recognition performance. To alleviate these problems, this paper proposes an new approach named ensemble convolutional echo state network, which takes Echo State Network (ESN) as the base classifier for ensemble and Convolutional Network (CN) to transform the input face image for further feeding to ESN, where the random parameters and architectures are assigned to ensure the diversity of the ensemble and to avoid computing stochastic gradient. Based on the rich dynamics of ESN and rich variations of input face image finished by CN, the proposed approach has the great ability to deal with the real facial expression recognition and to be scaled to the larger training data. It has also only one parameter to be adjusted. Conducted experiments show that the method achieves significant improvement over current methods on person- independent facial expression recognition.**

Keywords—facial expressions recognition; echo state network; convolutional network

I. INTRODUCTION

Facial expression is the most cogent means for humans to communicate emotions, to express intentions and to regulate interactions with the environment and the other humans [1]. It is reported that facial expression constitutes 55% of the effect of a communicated message while language and voice constitute 7% and 38% respectively [2], indicating that facial expression in the visual channel is the most effective and important cue that correlates well with the body and the voice[3]. Therefore, facial expression recognition (FER) has been widely studied so as to solve the practical problems such as to design the better consumer electronic devices [4]. The related methods can be broadly classified as two types: the single classifier and the ensemble classifier. The representatives of the first type includes neural network [5], SVM [1], sparse representation-based classification [6], metric learning-based methods [7]etc. The second type is based on the framework of ensemble learning[3][8][9]. In real environments, human facial appearance is easily influenced by age, gender, ethnicity, and possible occlusions due to eye-glasses and facial hair [10]. In such case, facial expressions tend to have overlapping features, making it difficult to find effective classification boundaries [7]. FER must be invariant or largely tolerant to these differences. In such case, ensemble learning is more suitable to deal with these problems. Ensemble learning has the ability to finish data fusion on heterogeneous features through training a single classifier on different feature sets independently. However, the performance of ensemble learning highly depends on the diversity of classifiers that make up the ensemble, requiring that individual classifier is unstable such as neural network [11].

Echo state network (ESN) is a newly developed method that has a dynamic reservoir containing a large number of sparsely interconnected neurons with non-trainable weights. It only has the output weights to be adjusted by the direct pseudo inverse calculation instead of gradient descent training, leading to the lower training computationally expensive [12][13]. Convolutional Network (CN) has been widely used for image recognition tasks, as it has the great ability to extract the powerful features [14]. This is very important, as the performance of a classifier is essentially decided by the represented facial expression features. A major barrier to the application of ESN and CN is that they currently require considerable skill and experience to choose sensible values for hyper-parameters such as the number of layers, the number of units per layer, the spectral radius and the degree of sparsity in the reservoir. The tuning of these parameters is especially expensive. Another barrier is that stochastic gradient descent used by them has been the challenge of scaling to the very large data sets and of learning very deep neural network models [14]. This is because the gradients tend to decrease, called “vanishing gradients" problem, when they are back-propagated through multiple levels of nonlinearity.

In order to make full use of ESN and CN while to avoid their disadvantages, this paper presents a novel approach for FER named as the ensemble convolutional echo state network (EC-ESN), which takes ESN as the base classifier for ensemble and CN to transform the input face image and then fed to ESN, where the random parameters and architectures are assigned to ESN and CN respectively so as to ensure the diversity of the ensemble and avoid to compute stochastic

This work was supported by China National Science Foundation under Grants 60973083, 61273363, State Key Laboratory of Brain and Cognitive Science under grants 08B12, Science and Technology Planning Project of Guangdong Province (2014A010103009)

gradient descent. The main contributions can be summarized as follows: ① To our best knowledge, it is the novel applications of ESN to FER, despite that it has been applied to speech emotion recognition [15]. ②It is a novel but effective idea to take ESN as the base classifier for ensemble, because ESN has a dynamic reservoir that can provide rich dynamics, which benefits for the diversity for an ensemble classifier. ③It is a novel idea to integrate CN with ESN, which allows the facial features under the large variation to be detected so as to provide the better input for ESN. ④EC-ESN is an innovative approach for FER, which integrates both ESN and CN with random parameters under the ensemble framework. EC-ESN can nicely deal with the larger training data, resisting in that the input face image with all kinds of rotation, translation and scaling.

II. Related Work

Our approach belongs to the ensemble learning, but provides the novel base classifier for the ensemble. Current there are three categories of ensemble approaches available for FER with attempts to provide the different individual classifier for the ensemble: ①The different training datasets are applied to train individual classifiers. Such datasets are often obtained through re-sampling techniques, such as bootstrapping or bagging, from the entire training data. To ensure that individual boundaries are adequately different, more unstable classifiers should be used as the base models, since they can generate sufficiently different decision boundaries even for small perturbations in their training data. ②The individual classifiers can be achieved by using different features, or different subsets of existing features[3][8] [16]. For example, the support vector machine as the base classifier is created using gabor filters and local binary patterns [8]. An ensemble of features, scale-invariant feature transform, and some coarse motion features has been applied to facial expression recognition where support vector machines is taken as the classifier [3]. ③The different training parameters are used to generate the base classifiers. Adjusting such parameters allows one to control the instability of the individual classifiers and hence contribute to their diversity. For example, a series of multilayer perceptron neural networks can be trained by such as different weight initializations and number of layers. The other examples involve in taking an extreme learning machine as the base classifier [17]. Our approach results from the ideas of last two categories, which applies ESN with random parameters to generate the different classifier and applies CN with random parameters to generate the different feature sets. Despite ESN has been applied to design an automatic speech emotion recognition system [15], it is primarily applied to time series prediction [18], instead of classification. CN with multilayer neural networks have recently made significant progress in a variety of image classification and detection tasks [19][20]. It requires a much large data sets for training whereas stochastic gradient descent is applied to learning the parameters. This is the much challenge of learning very deep neural network models. Different from above work, our approach does not require training and the parameters are assigned at random, so that the trouble is avoided. To our knowledge, the integration of CN and ESN has not been found in the current literatures.

III. Preliminaries

A. Ensemble Learning

An ensemble of classifiers consisted of a combination of different homogeneous or heterogeneous classifiers aims to jointly perform a classification task. A classifier is a function $f: X \rightarrow Y$ that maps an instance $x \in X \subseteq R^D$ onto an element of Y. The difficulty is to find a definition for the unknown function *f(x)*. In the case of ensembles, givn a set of base classifiers $C = \{C_1, C_2, \cdots, C_m\}$, each of which map $x \in R^D$ onto a label in the set $Y = \{1, \cdots, K\}$, an ensemble is constructed by solving two main problems: the construction of the individual classifiers C_i and the design of the combination rule that allows the ensemble to achieve the label of the instance x from the outputs of the individual classifiers $\{C_1(x), C_2(x), \cdots, C_m(x)\}$. The combination rules include the simple majority voting, weighted majority voting, a separate classifier for stacking, etc. In our approach, the simple voting is applied.

B. Echo State Network

ESN consists of a non-trainable reservoir and a linear readout part [12][13]. The reservoir can maintain active even in the absence of input and exhibits dynamic memory. Connection weights inside the reservoir and input weights are generated randomly, while only the connections between the reservoir and output readout neurons are adaptable via supervised learning. Therefore training an ESN network becomes a simple linear regression task, which expedites convergence.

Consider a discrete time ESN with K input units, N internal units and L outputs, as depicted in Fig. 1. The input units, specified by the vector $u(n)=[u_1(n), \cdots, u_K(n)]^T$, are linearly combined, according to the weights contained in matrix $\boldsymbol{W}^{in}$, and transmitted to the internal reservoir, which is composed of fully connected nonlinear neurons whose activations, given by $x(n)=[x_1(n), \cdots, x_N(n)]^T$, represent the network states, and are updated by:

$$x(n+1) = f(W^{in}u(n+1) + Wx(n) + W^{back}y(n)) \quad (1)$$

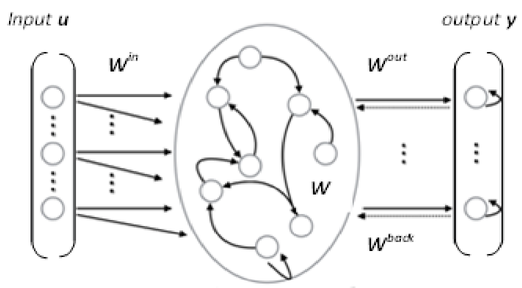

Fig. 1. Basic architecture of ESN

where $\boldsymbol{W}$ contains the weights of the connections within the reservoir, $\boldsymbol{W}^{back}$ specifies the weights for the connections that send the network outputs back to the reservoir layer, and $f=(f_1,\cdots,f_N)$ denotes the activation functions of the internal units. Finally, the network outputs, represented as the vector $y(n)=[y_1(n),\cdots,y_L(n)]^T$, are determined according to the following formula:

$$y(n+1) = f^{out}(W^{out}x(n+1)) \quad (2)$$

where f^{out} specifies the activation functions of the output units, which usually corresponds to the identity function, so that the output layer becomes a linear combiner. In our approach, $\boldsymbol{W}$, $\boldsymbol{W}^{in}$, $\boldsymbol{W}^{back}$ are determined at random before learning, whereas $\boldsymbol{W}^{out}$ is determined by the supervised learning.

C. Convolutional Networks

CN contains convolutional layers and sub-sampling layers. Each convolutional layer is interspersed with sub-sampling layer to reduce computation time and to gradually build up further spatial and configural invariance. At a convolution layer, the previous layer's feature maps are convolved with learnable kernels and put through the activation function to form the output feature map. Each output map may combine convolutions with multiple input maps, such as by all-pairs or all triplets. For a particular output map, the input maps would be convolved with distinct kernels. At a sub-sampling layer, it produces down-sampled versions of the input maps. If there are M input maps, there will be exactly M output maps, although the output maps will be smaller. Typically sub-sampling function will sum over each distinct n-by-n block in the input image so that the output image is n-times smaller along both spatial dimensions.

IV. ENSEMBLE CONVOLUTIONAL ECHO STATE NETWORK

A. The Overall Framework

Fig. 2 presents the framework of EC-ESN that consists of two main parts: CESN (Convolutional Echo State Network) as the base classifier for the ensemble and fusion is performed by the simple majority voting strategy for the final decision.

In EC-ESN, the entire face image is fed to each CESN which classifies the facial expression of the input face image into one of the seven coded emotional state vectors. For example, the emotion state of neutrality is coded as a vector

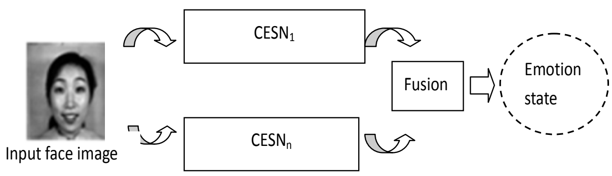

Fig. 2. The overall framework of EC-ESN

v=(1 0 0 0 0 0 0) while the emotion state of surprise is coded as a vector v=(0 0 0 0 0 0 1). CESN is the base classifier composed of CN and ESN, shown as Fig. 3, where CN extracts rich features for the input face image through convolutional transformation and then fed to ESN to perform the classification.

CN is exclusively selected here to contain the variants of the input face image so as to enhance the generalization of CESN. However, the performance is still determined by multiple interrelated architectural factors, such as the number of layers, feature map dimensions, the number of parameters, and pooling sizes. These factors must be carefully designed using good intuition along with extensive trial-and-error experiments on a validation set, as they are conflated by the capacity of the model and the degree of non-linearity. For example, adding another layer increases the number of parameters, but it also puts an additional non-linearity into the system. Instead of carefully selecting the appropriate values for these factors, we random assign the values for them within an ensemble framework. This method can also avoid the problem of computing stochastic gradient descent for deep layers.

ESN has the rich dynamics for its dynamic reservoir decided by a few parameters such as the spectral radius and the degree of sparsity. The dynamic reservoir is responsible for generating a set of signals that will be linearly combined to produce the outputs. It is much attractive to create a rich set of dynamics in order to reach an adequate approximation of the desired dynamic behaviors. Intuitively, a sparse connection structure within the reservoir tends to decouple groups of neurons, which contributes to the development of individual dynamics, i.e., distinct and lowly correlated internal states. It can be inferred that the dynamic reservoir with much different parameters will result in much different dynamics. This is why ESN with random parameters can be taken as the base classifiers to ensure the diversity for the ensemble.

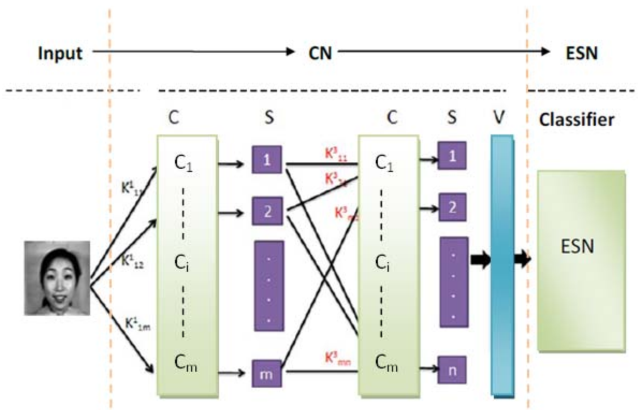

Fig. 3. The architecture of CESN

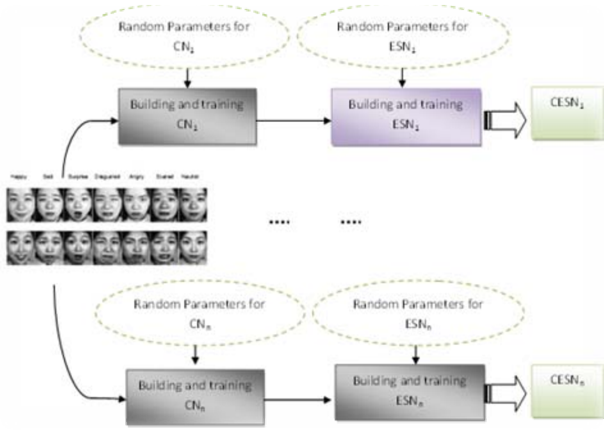

Fig. 4. Generating multiple CESNs as the base classifiers

B. Algorithm

Instead of finding the optimal training parameters that may require testing hundreds of parameter values, each base classifier CESN for the ensemble is generated by assigning random parameters to CN and ESN respectively, shown as Fig.4, where total n individual CESNs are generated.

EC-ESN is summarized as a two stages process. In the first stage all individual classifiers are constructed. The voting method proceeds to the stage two to make classification for the query facial expression.

Algorithm 1. EC-ESN
Input: Input facial expression images training data $\boldsymbol{D}=\{(x_i,y_i)\}$, the query facial expression image x, and the number of ensemble classifiers n, where y_i be the class of x_i, belong to a set of classes$\{C_1,..,C_T\}$
Output:the emotion label y for x

Stage 1: training
[1] For i=1:n
  a) Generate random parameters to construct CN_i
  b) Generate random parameters matrices ($\boldsymbol{W^{in}},\boldsymbol{W}, \boldsymbol{W^{back}}$), and apply them to generate ESN_i
  c) Apply the training data $\boldsymbol{D}$ to CN_i to generate the output X_{cn}
  d) Apply X_{cn} to train ESN_i and achieve the final individual classifier, denoted as $CESN_i$
[2] END
Stage 2: predicting
[3] Apply each $CESN_i$ to classify x. Let $v_{i,j}=1$ if $CESN_i$ classify x to class C_j, or else $v_{i,j}=0$
[4] Obtain the total vote received by each class, $\mathrm{V}_j=\sum_{i=1}^{T} v_{i,j}, j=1\cdots T$
[5] Choose the class assigned to y for x that receives the highest total vote

---

V. Experimental Results

Experiments are conducted to validate the proposed approach on JAFFE and Cohn-Kanade (CK) databases that have been widely used in elsewhere [7][8][17]. Some state-of-the-art techniques with related to our approach are also compared.

A. Databases

JAFFE has 213 facial images in total and ten expressers posed seven kinds of expressions (neutral, happy, sadness, surprise, anger, disgust and fear). For each individual there are 3-4 sample images, each of which is 256 x 256 pixels. The frontal facial images in the database are nearly the same size and the hair was tied away from the face to expose all the expression zones, but there are differences in light intensity. CK consists of 593 sequences taken from 123 subjects. The sequences represent a set of facial expressions performed by individuals, each of which is 640 x 490 pixels. Here only the first and the last two frames in each sequence are applied. In experiments, the images from two databases are used directly without using any feature extraction method. Besides, each image is resized into 64 x 64 pixels. It is expected that the method can learn the features useful to make classification, reducing the human load to extract and select features.

B. Experimental Setting up

In experiments, two strategies are applied: person-dependent (PD) and person-independent (PI) [7]. In PD strategy, training expression images and testing expression images are selected randomly from the whole database, where the expression of the same persons may appear both in the training set and the test set. In PI strategy, the persons are randomly selected and make sure that the same person will not appear in the training set and test set simultaneously. The baseline methods selected to make comparison are SVM, SRC, Softmax, and ESN, as they are most representative and successful in facial expression recognition. The parameters for each method are assigned respectively using the cross validation on the training data. For SVM, rbf kernel is utilized, C is selected from $\{2^{-2},2^{-1},..,2^{12}\}$, and gamm is selected from $\{2^{-10},2^{-9},..,2^{4}\}$. SRC has not any parameter to be selected. For Softmax, its iteration number takes 300. ESN selects the number of neurons in reservoir from {100,200,..,1000}, and selects the spectral radius and the degree of sparsity from {0.1,0.2,..,1.0} respectively. In all experiments, the accuracy is employed to evaluate all the approaches.

C. Property Analysis for EC-ESN

EC-ESN has only one parameter to be adjustable, that is, the number of individual classifiers for the ensemble. Choosing the ensemble size involves balancing speed and accuracy, as the larger ensemble takes longer to train and to generate predictions. Generally ensemble methods can become over-trained when the ensemble size is too large, but a smaller ensemble size always cannot reach the ideal accuracy. Hence, an appropriate size needs to be setup carefully but expensively.

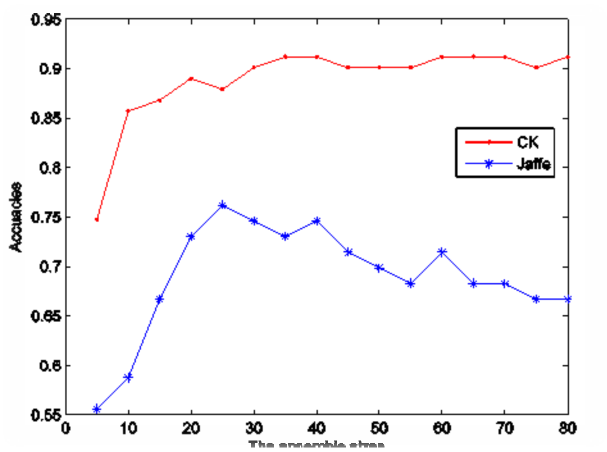

Fig. 5. Accuracies of EC-ESN against the ensemble sizes

It is expected that the classifier is robust to the ensemble size so as to avoid over-trained problem. It can be observed from Fig. 5 that EC-ESN performs better by taking the larger ensemble sizes on CK, while taking the smaller ensemble sizes on JAFFE, illustrating that the larger the training data is, the bigger the ensemble size should take. This can be easily understood, as the diversity of the ensemble on the smaller training data cannot be ensured if the larger ensemble size is taken. It can be also observed that the overall trend of accuracies varying with the ensemble sizes is stable, showing that the optimal ensemble size can be determined easily. For example, it can be observed that the ensemble size of EC-ESN should take the value of 80 on the CK and take the value of 25 on the JAFFE, around which the accuracies do not vary drastically. On the other hand, most methods are confronted with the over-trained problem. To illustrate that EC-ESN has ability to deal with the problem, the experiments with PI are performed by varying the data size on CK database.

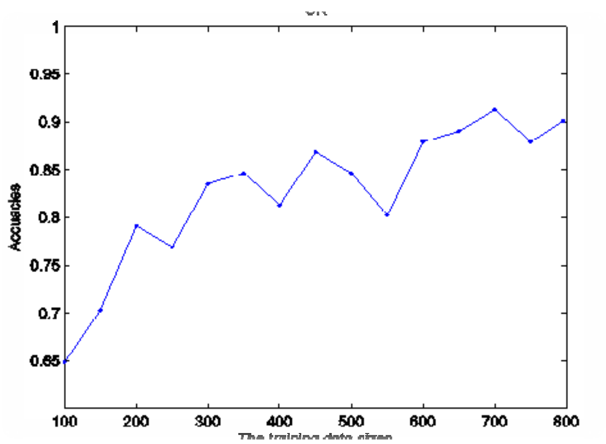

Fig. 6. Accuracies of EC-ESN against the training data sizes

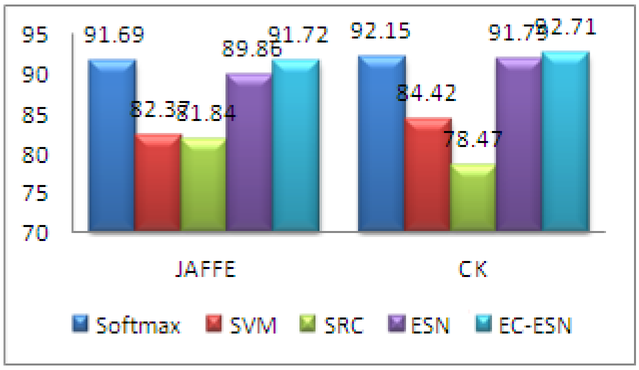

Fig. 7. Accuacies (%) of all methods by PD on two databases

It can be seen from Fig.6 that the accuracies of EC-ESN increases with the larger data on the whole, illustrating that EC-ESN can be nicely scaled to the larger training data.

D. *Performance Comparison*

The experiments are firstly conducted with PD strategy to validate EC-ESN, where five-fold cross is applied. The average accuracy is obtained by averaging the accuracies of five trials. It can be observed from Fig. 7 that EC-ESN works well as the same as Softmax on both databases, up to 91% on JAFFE and 92% on CK respectively. However, EC-ESN does not significantly outperform the baseline methods. This is because in such cases ESC as the base classifier for ensemble is too strong whereas the diversity is not ensured. On the other hand, the experiments with PI are also conducted. It can be seen from Fig. 8 that EC-ESN performs much better than the compared approaches. EC-ESC outperforms Softmax about 13% on JAFFE and about 10% on CK, where Softmax is the second top one in terms of the accuracy. This improvement is significant, showing that EC-ESN has great potential ability to recognize expressions of new faces in real environments. Notably, SVM and SRC perform very badly in this case, possibly due to that they require the carefully selected features.

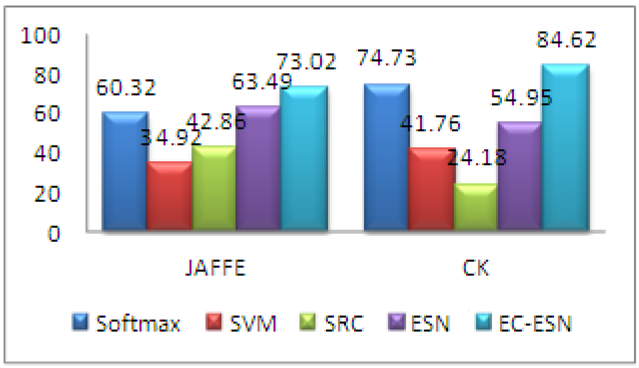

Fig. 8. Accuacies (%) of all methods by PI on two databases

TABLE I. TIME FOR PREDICTING EACH FACIAL EXPRESSION (SECONDS)

	Softmax	SVM	SRC	ESN	EC-ESN
JAFFE	0.0015	0.0171	2.5803	0.1540	0.9324
CK	0.0012	0.0427	18.813	0.1527	0.5993

From experimental results with PD and PI, it can be found that facial expression is usually correlated with identity and variations in identity dominate over those expressions. This is consistent with the results that most of the compared methods including Softmax seem to perform well on expression recognition by PD, but are substantially less efficient on the expression recognition by PI. Particularly, in the real applications, it is difficult to collect the different expressions of the user as the training set. In such case, our approach obviously stood out for the real environment.

E. Time Comparison

The predicting time for the query facial expression is much critical. The predicting times taken by each method are presented in Table1, where the predicting time of EC-ESN has been divided by the ensemble size, as it can be implemented in parallel. It can be observed that Softmax predicts the facial expression much more quickly than the other methods. EC-ESN can perform the facial expression recognition in less than one second, fitting for real applications. SRC is seen to be most costly one among these methods and cannot be scaled to the large training data, as it is the lazy learning method. The training times are not compared, as the training for each method can be finished off-line. Besides, each method involves in multiple parameters that have much influence on their training times, leading to the unfair comparison.

VI. CONCLUSION

This paper presents a novel ensemble approach for FER, which is integrated with echo state network and conventional network within the framework of ensemble learning. The proposed approach can alleviate the possible overlapping features of different facial expressions, as conventional faces from the original face represents a lot of variants of the face. It also has one parameter only to be adjustable so as to make it more practicable. Finally it can be nicely scaled to the larger data with the better performance, as CN and ESN can be adaptable to the large train data. Like any ensemble method, the training and testing phase of our approach are also time consuming tasks. These problems will be solved with parallel computing technique[21], such that our approach in the recognition phase can run at real-time. On the other hand, the fusion in our approach is performed by the simple voting strategy and the detection of diversity has not been considered. These two problems will be investigated so as to further enhance the performance of our approach in the future.

REFERENCES

[1] Niki Aifanti, Anastasios Delopoulos, "Linear subspaces for facial expression recognition", Signal Processing: Image Communication 29 (2014)177–188

[2] Amal Azazi, Syaheerah Lebai Lutfi, Ibrahim Venkat, Fernando Fernandez-Martinez, "Towards a robust affect recognition: Automatic facial expression recognition in 3D faces", Expert Systems with Applications 42 (2015) 3056–3066

[3] Tariq, U., Kai-Hsiang Lin, Zhen Li, Xi Zhou and Zhaowen Wang, "Recognizing Emotions From an Ensemble of Features", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Volume: 42 , Issue: 4,2012 , pp.1017 – 1026.

[4] Bernard Fong, Joyce Westerink, "Affective Computing in Consumer Electronics", IEEE transactions on affective computing, 3(2),2012

[5] Chathura R, De Silva, Surendra Ranganath, Liyanage C. De Silva, "Cloud basis function neural network: a modified RBF network architecture for holistic facial expression recognition", Pattern Recognition 41 (2008) 1241 – 1253

[6] Chen Ying, Zhang Shiqing, Zhao Xiaoming, "Facial expression recognition via non-negative least-squares sparse coding," Information (Switzerland), v 5, n 2, pp. 305-318, 2014

[7] Shaohua Wan and J.K.Aggarwal, "Spontaneous facial expression recognition: A robust metric learning approach", Pattern Recognition 47(2014)1859–1868

[8] Zavaschi, Thiago H.H., Britto Jr., Alceu S., Oliveira, Luiz E.S.,Koerich, Alessandro L., "Fusion of feature sets and classifiers for facial expression recognition", Expert Systems with Applications, v 40, n 2, p 646-655, February 1, 2013

[9] Jia Xibin , Zhang Yanhua, Powers David, Ali, Humayra Binte , "Multi-classifier fusion based facial expression recognition approach", KSII Transactions on Internet and Information Systems, v 8, n 1, pp.196-212, 2014

[10] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, Lijun Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey", Image and Vision Computing 30 (2012) 683–697

[11] Stefan Faußer,Friedhelm Schwenker, "Neural Network Ensembles in Reinforcement Learning", Neural Process Lett (2015) 41:55–69

[12] Rodan A, Tino P. "Minimum Complexity Echo State Network". IEEE Transactions on Neural Networks, 2011, 22(1): 131-144.

[13] Han, Seong I. Korea , Jang M. "Fuzzy Echo State Neural Networks and Funnel Dynamic Surface Control for Prescribed Performance of a Nonlinear Dynamic System", IEEE Transactions on Industrial Electronics, v 61, n 2, pp. 1099-1112, 2014

[14] David Eigen, Jason Rolfe, Rob Fergus and Yann LeCun, "Understanding Deep Architectures using a Recursive Convolutional Network" , International Conference on Learning Representations (ICLR2014), 2014

[15] Stefan Scherer, Mohamed Oubbati, Friedhelm Schwenker, and G ¨ unther Palm, "Real-Time Emotion Recognition from Speech Using Echo State Networks", L. Prevost, S. Marinai, and F. Schwenker (Eds.): ANNPR 2008, LNAI 5064, pp. 205–216, 2008.

[16] Nicolás García-Pedrajas, Jesús Maudes-Raedo, César García-Osorio, Juan J. Rodríguez-Díez, "Supervised subspace projections for constructing ensembles of classifiers", Information Sciences, Volume 193, 15 June 2012, pp.1-21

[17] Ghimire, Deepak; Lee, Joonwhoan, "extreme learning machine ensemble using bagging for facial expression recognition", Journal of Information Processing Systems, v 10, n 3,pp. 443-458, 2014

[18] Min Han and Meiling Xu, "Subspace Echo State Network for Multivariate Time Series Prediction", T. Huang et al. (Eds.): ICONIP 2012, Part V, LNCS 7667, pp. 681–688, 2012

[19] Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Michaël Mathieu and Yann LeCun, "Learning Convolutional Feature Hierachies for Visual Recognition", Advances in Neural Information Processing Systems (NIPS 2010), 23, 2010

[20] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks",In Advances in Neural Information Processing Systems 25 (NIPS'2012), 2012.

[21] Carlos Valle,Francisco Saravia , Héctor Allende, Raúl Monge, César Fernández, "Parallel Approach for Ensemble Learning with Locally Coupled Neural Networks", Neural Process Lett(2010) 32:277–291