Journal of Information Hiding and Multimedia Signal Processing ©2013 ISSN 2073-4212
Ubiquitous International Volume **4**, Number **3**, July **2013**

Semi-supervised Facial Expression Recognition Algorithm on The Condition of Multi-pose

Bin Jiang and Kebin Jia

School of Electronic Information & Control Engineering
Beijing University of Technology
No.100, Pingleyuan, Chaoyang District, Beijing 100124, China
bj2009@emails.bjut.edu.cn; kebinj@bjut.edu.cn

Received December, 2012; revised February, 2013

Abstract. *A major challenge in pattern recognition is labeling of large numbers of samples. This problem has been solved by extending supervised learning to semi-supervised learning. Thus semi-supervised learning has become one of the most important methods on the research of facial expression recognition. Frontal and un-occluded face images have been well recognized using traditional facial expression recognition based on semi-supervised learning. However, pose-variants caused by body movement, may decrease facial expression recognition rate. A novel facial expression recognition algorithm based on semi-supervised learning is proposed to improve the robustness in multi-pose facial expression recognition. In the proposed method, transfer learning has been brought into semi-supervised learning to solve the problem of multi-pose facial expression recognition. Experiments show that our method is competent for semi-supervised facial expression recognition on the condition of multi-pose. The recognition rates are 82.68% and 87.71% on the RaFD database and BHU database, respectively.*

Keywords: facial expression recognition; semi-supervised learning; muti-pose; transfer learning

1. **Introduction.** As the most important part of human-computer interaction and affective computing, facial expression recognition relies to a large extend on number of labeled images for training classifier. Large sample can better reflect the real distribution of samples and consequently obtaining good generalization error. Image labeling, however, is very complex, expensive and time consuming, as it requires the efforts of experienced human annotators. Semi-supervised learning [1] addresses this problem by using both labeled data and large amount of unlabeled data, to build better classifiers. In 2004, Cohen et al [2] proposed a facial expression recognition method based on Bayesian networks using both labeled training data and plenty of unlabeled training data. They applied semi-supervised learning on facial expression recognition firstly. In 2009, Hady et al [3] proposed a learning framework to exploit the unlabeled data by decomposing multi-class problems into a set of binary problems and applying Co-Training to solve each binary problem. The results show that their method improves the recognition accuracy of facial expressions. In 2011, Chen and Wang [4] took all three semi-supervised assumptions including smoothness, cluster, and manifold assumptions, into account during boosting learning. Experiment demonstrates that their algorithms yield better results for facial expression recognition tasks in comparison to other state-of-the-art semi-supervised learning algorithms.

Traditional semi-supervised learning methods in face recognition and facial expression recognition use the frontal and un-occluded face images for experiments. However, these pre-conditions are not always satisfied in most real-world applications. The performance of current facial expression recognition systems may decrease significantly on multi-pose case. Pose variation is one of the restrictions in automatic facial expression recognition. The major challenge of the pose problem is that face image may be squashed or stretched and human face may be occluded as well. Therefore pose variation will jeopardize the integrity of facial expressions. Even if two samples have the same facial expression, feature distribution of them is also different. Nevertheless, most of semi-supervised learning methods assume that the training and test data must be in the same feature space and have the same distribution. In multi-pose facial expression recognition application, this assumption may not be fulfilled.

In recent two years, some supervised learning methods were proposed for multi-pose face recognition problem. Li et al [5] discussed the single training sample based face recognition under variable face poses. Li and Chen [6] proposed a method of multi-view face recognition based on pose estimation combining tensor face and manifold learning effectively. Wang et al [7] used orthogonal discriminant vector to recognize face images across pose. Li et al [8] considered that subspace-based face representation can be looked as a regression problem. The approach for cross-pose face recognition was proposed by using a regressor with a coupled bias-variance tradeoff. Lee et al [9] decomposed the appearance of the face in each pose class using an embedded hidden markov model, the method had been proven to be effective in the recognition of faces across poses. Even though these methods improve the robustness and performance of face recognition, facial expression is a more challenge task. In addition semi-supervised learning is employed to improve the robustness of facial expression recognition.

Transfer learning [10] theory may offer a way to improve the semi-supervised learning. According to knowledge transfer, the theory allows the domains, tasks, and distributions used in training and testing to be different. Dai et al [11] proposed a boosting based algorithm named TrAdaBoost, which is an extension of the AdaBoost [12] algorithm. TrAdaBoost assumes that, due to the different distributions between training and test data, some of training data may be useful in learning for test data but some of them may not and could even be harmful. It attempts to iteratively reweight training data to reduce the weight of the bad training data while encourage the good training data to contribute more for test data classification. So based on TrAdaBoost algorithm, we propose a novel algorithm called Multi-Pose Adaptive Boosting (MP-AdaBoost). MP-AdaBoost employs the knowledge transfer to construct a high-quality classification model for multi-pose facial expression data. Our experimental results show that MP-AdaBoost is an effective and robust method.

The rest of the paper is organized as follows. In Section 2, we present the principle of MP-AdaBoost. Our experimental results and discussion are shown in Section 3. Section 4 concludes the whole paper.

2. MP-AdaBoost Algorithm.

2.1. TrAdaBoost Algorithm.

Let X_d be the different-distribution feature space, X_s be the same-distribution feature space, and $Y = \{0, 1\}$ be the set of category labels. A concept is a boolean function c mapping from X to Y, where $X = X_s \bigcup X_d, x_i \in X, i = 1, 2, \ldots, n + m$. The first n samples belong to X_d, the rest of X belong to X_s.

TrAdaBoost is based on attenuation to transfer the knowledge about X_d into training dataset X and construct classification model for testing dataset S. TrAdaBoost is described as follows:

Initialize the initial weight vector, that $W^1 = (w_1^1, w_2^1, \ldots, w_{n+m}^1)$.

For $t = 1, 2, \ldots, N$

1. Set

$$P^t = W^t / (\sum_{i=1}^{n+m} w_i^t) \tag{1}$$

2. Call Learner to classify dataset with the distribution P^t. Then, get back a hypothesis $h_t : X \to Y, Y = \{0, 1\}$.

3. Calculate the error of h_t on X_s:

$$\epsilon_t = \sum_{i=n+1}^{n+m} \frac{w_i^t |h_t(x_i) - c(x_i)|}{\sum_{i=n+1}^{n+m} w_i^t} \tag{2}$$

4. Set $\beta_t = \epsilon_t/(1-\epsilon_t), \beta = 1/(1+\sqrt{2 \ln n/N})$. Note that, ϵ_t is required to be less than 0.5.

5. Update the new weight vector:

$$w_i^{t+1} = \begin{cases} w_i^t \times \beta^{|h_t(x_i)-c(x_i)|}, & i = 1, 2, \ldots, n \\ w_i^t \times \beta^{-|h_t(x_i)-c(x_i)|}, & i = n+1, n+2, \ldots, n+m \end{cases}$$

Output the hypothesis:

$$h_f(x) = \begin{cases} 1, \prod_{t=\lceil N/2 \rceil}^{N} \beta_t^{-h_t(x)} \geq \prod_{t=\lceil N/2 \rceil}^{N} \beta_t^{-\frac{1}{2}} \\ \\ 0, otherwise \end{cases} \tag{3}$$

The framework of TrAdaBoost is shown that AdaBoost is similar to TrAdaBoost in many respects. Both of them aim to boost the accuracy of a weak learner by carefully adjusting the weights of training data and learn a classifier accordingly. But AdaBoost don't consider whether distributions of the training and test samples are identical, this algorithm is mainly careful for decreasing classification error. TrAdaBoost has divided training dataset into two parts, X_s and X_d. Because X_s and S are from the same distribution, when working on X_s, TrAdaBoost is equivalent to AdaBoost. However, X_d and S are from the different distribution, when they are wrongly predicted due to distribution changes by the learned model, TrAdaBoost will add a mechanism to decrease the weights of these training samples in order to weaken their impacts.

Every training sample of X_d is multiplying its training weight by $\beta^{|h_t(x_i)-c(x_i)|}$ in each iteration round. Note that $\beta^{|h_t(x_i)-c(x_i)|} \in (0, 1]$ If the classification result is correct, training weight will be the same weight as in the last iteration. If the sample is wrong predicted, its training weight will be reduced. After several iterations, the samples in the X_d that well fit the same-distribution ones, will have larger training weights, while the other in the X_d that are dissimilar to the same-distribution ones will have lower weights. The samples with large training weights will be more helpful to train better classifiers. In extreme cases, if none of X_d is fitting the same-distribution ones, TrAdaBoost can't transfer knowledge. In turn the algorithm will be equivalent to AdaBoost.

2.2. **MP-AdaBoost Algorithm.** MP-AdaBoost has still used the knowledge transfer and improved TrAdaBoost in two aspects:

Firstly, in order to solve the problem of semi-supervised facial expression recognition, MP-AdaBoost is proposed which uses large amount of unlabeled data together with the labeled data to build classifiers. As mentioned above, some randomly chosen training samples for one expression become labeled data. The rest of training samples become unlabeled data. Applying nearest neighbor method, we can get their pseudo labels. Thus MP-AdaBoost not only meets the requirements of semi-supervised learning, but also retains the ability of knowledge transfer.

Secondly, TrAdaBoost is an extension of the AdaBoost. Both of them can only deal with two-class problem. However, facial expression recognition is a multi-classification problem. The improved MP-AdaBoost can recognize multi-class faical expressions. In extreme cases, if none of X_d is fitting the same-distribution ones and pseudo-classes are same to true labels, MP-AdaBoost will be equivalent to AdaBoost.M1 [12] algorithm, which is multi-class AdaBoost algorithm.

MP-AdaBoost is given as follows:

Nearest neighbor assignment is used to assign the initial pseudo-class labels and the initial weight vector is initialized that

$$W^1 = \begin{cases} w_i^1 = 0.9/[(n+m)/2], w_i^1 \text{ represents labeled data} \\ \\ w_i^1 = 0.1/[(n+m)/2], w_i^1 \text{ represents unlabeled data} \end{cases}$$

For $t = 1, 2, \ldots, N$

1. Set $P^t = W^t/(\sum_{i=1}^{n+m} w_i^t)$

2. Call Learner to classify dataset with the distribution P_t. Then, get back a k-class hypothesis: $h_t : X \to Y, Y = \{1, 2, \ldots, k\}$.

3. Calculate the error of h_t on X_s:

$$\epsilon_t = \sum_{i=n+1}^{n+m} \frac{w_i^t e_i^t}{\sum_{i=n+1}^{n+m} w_i^t} \tag{4}$$

where e_i^t represents the multi-classification result of i-th training sample at round t. If training sample is wrong predicted, $e_i^t = 1$. If training sample is correctly predicted, $e_i^t = 0$.

4. Set $\beta_t = \epsilon_t/(1-\epsilon_t), \beta = 1/(1+\sqrt{2 \ln n/N})$. Note that, ϵ_t is constrained to be less than 0.5.

5. Update the new weight vector:

$$w_i^{t+1} = \begin{cases} w_i^t \times \beta^{e_i^t}, & i = 1, 2, \ldots, n \\ w_i^t \times \beta^{-e_i^t}, & i = n+1, n+2, \ldots, n+m \end{cases}$$

Output the hypothesis:

$$h_k(x_i) = argmax \sum_{t=1}^{N} (\log \frac{1}{\beta_t})(1 - e_i^t) \tag{5}$$

The MP-AdaBoost algorithm practically inherits all the good properties of TrAdaBoost. In step 5, we can see that if the samples from different-distribution are wrong predicted, their weights will be reduced. On the other hand, MP-AdaBoost algorithm can achieve multi-classification and semi-supervised classification. In step 3, e_i^t is helpful to solve the multi-classification problem. Equation (5) is inherited from AdaBoost.M1 algorithm. With the steps above, MP-AdaBoost algorithm can achieve semi-supervised facial expression recognition algorithm on the multi-pose condition.

3. Experimental Results.

3.1. Experiment Procedure. In order to test the robustness and effectiveness of the proposed MP-AdaBoost algorithm, Radboud Faces Database [13] (RaFD) and Beihang University facial expression database [14] (BHU) are chosen for the experiment. The RaFD is a high quality database, which contains the out-of-plane rotation facial expression images including 0 degree, 45 degree, frontal face, 135 degree and 180 degree, from the right side to the left side. The BHU database contains the frontal and 30 degree profile of each facial expression of every subject. Some samples of two databases are shown in Fig 1 and Fig 2.

Figure 1. Samples of RaFD Database

Figure 2. Samples of BHU Database

Our experimental procedure is divided into three stages: preprocessing, feature extraction and expression classification. Firstly, we crop the original image of database into 64×64 by removing the back-ground influences. Since the illumination condition is varied to the images in both databases, we apply histogram equalization to eliminate lighting effects. Secondly, linear discriminant analysis (LDA) is used to extract important features from each image. For the facial expression samples, the goal of LDA algorithm is to maximize the between-class measure while minimizing the within-class measure. Finally, we adopt the baseline algorithms (AdaBoost.M1, Label Propagation [15], ASSEMBLE [16], RegBoost [17]) and MP-AdaBoost to identify six facial expressions (angry, disgust, sad, happy, fear and surprise) in the RaFD and BHU databases. Label Propagation is a classical Semi-Supervised Learning method. AdaBoost.M1, ASSEMBLE and RegBoost are based on the framework of AdaBoost, so k nearest neighbor (k-NN) or back propagation neural network (BP-NN) becomes the basic classifier of experimental methods. The number of basic classifier and iteration is ten.

3.2. Experimental Results and Performance Comparisons.

3.2.1. *Experiments on RaFD database.* In order to test the performance of the proposed MP-AdaBoost algorithm in this paper, 1026 face images are chosen. The face images contain six expressions captured on three angles, the frontal face images are the samples X_d, 135 degree or 180 degree face images are the samples S. We randomly select some samples of S per expression as X_s, together with all samples of X_d to comprise training dataset. The number of chosen samples from S is proportional (10% to 50%) to the image quantity in X_d. The rest images in S are used as test samples.

In terms of different samples, we repeat the same procedure of training and testing for ten times. At last, we average all the ten recognition rates to obtain the final performance of the proposed algorithm. Note that, according to the requirement of semi-supervised learning, the number of labeled and unlabeled images in training dataset is 1:1. Some experimental samples are shown in Fig 3.

FIGURE 3. Experimental Samples of RaFD Database

Table 1 shows the performance comparisons among our proposed algorithm and baseline algorithms. The experimental samples are selected from the frontal face and 135 degree face images. k-NN is the basic classifier, and k is equal to three.

TABLE 1. Comparison of classification accuracies on frontal and 135 degree face images

	10%	20%	30%	40%	50%
AdaBoost.M1	59.15%	65.65%	66.67%	70.49%	70.89%
LP	45.49%	54.67%	55.13%	60.39%	60.54%
ASSEMBLE	59.58%	66.27%	67.21%	71.08%	71.49%
RegBoost	59.54%	66.27%	67.29%	70.98%	71.61%
MP-AdaBoost	59.61%	66.41%	67.46%	71.18%	71.31%

The experimental samples in Table 2 are comprised by the frontal face and 180 degree face images. BP-NN is the basic classifier. From the results we can see that MP-AdaBoost algorithm still outperforms the other algorithms.

There are two major issues in the experiments: the first one is that the distribution of frontal face images is different from 135 degree or 180 degree face images. So if we can't find more facial expression information from frontal face images, where the number of samples is small, the classification accuracy may be very low; the other problem is that out-of-plane rotation often occludes face components and consequently much useful

Table 2. Comparison of classification accuracies on frontal and 180 degree face images

	10%	20%	30%	40%	50%
AdaBoost.M1	61.90%	71.45%	73.08%	75.34%	75.65%
LP	40.46%	48.51%	50.79%	53.38%	53.10%
ASSEMBLE	64.84%	68.19%	65.67%	72.25%	73.39%
RegBoost	65.78%	70.11%	72.25%	70.05%	65.71%
MP-AdaBoost	72.22%	76.34%	73.58%	80.20%	82.68%

expression information are lost. Both issues have negative impact on semi-supervised classification. But we can see that MP-AdaBoost algorithm has fulfilled semi-supervised learning. Furthermore, our proposed algorithm has transferred expression knowledge from frontal face images to out-of-plane rotation face images and solved the difficulties to some extent. From the aspect of accuracy, MP-AdaBoost can outperform baseline algorithms based on RaFD database.

3.2.2. *Experiments on BHU database.* Each sequence of BHU database shows a humans expression. We cut ten frames about peak expression from each sequence. Experiments choose 960 images from 96 videos, which contains six expressions and two angles (frontal face and 30 degree). The experimental setting is the same to 3.2.1. The basic classifier is k-NN, and k is equal to three. Some experimental samples are shown in Fig 4.

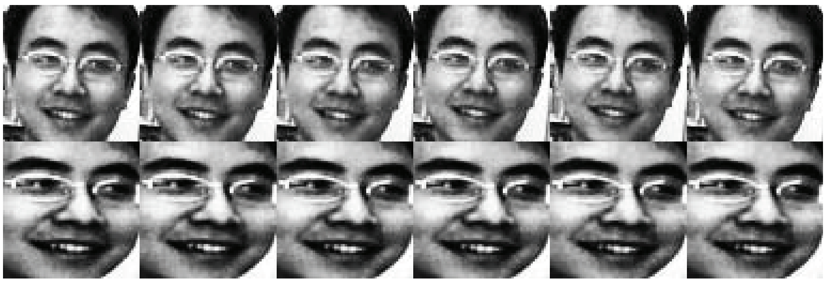

Figure 4. Experimental Samples of BHU Database

Table 3. Comparison of classification accuracies on frontal and 30 degree face images

	10%	20%	30%	40%	50%
AdaBoost.M1	51.81%	66.07%	76.07%	82.95%	87.54%
LP	33.63%	42.37%	46.34%	54.03%	63.83%
ASSEMBLE	52.06%	66.88%	76.58%	83.19%	87.67%
RegBoost	52.29%	66.54%	76.37%	83.33%	87.67%
MP-AdaBoost	52.31%	67.03%	76.31%	83.40%	87.71%

Table 3 shows that MP-AdaBoost is still superior to baseline methods in most cases. The experimental results show that our proposed algorithm is robust and capable to identify different facial expressions efficiently.

4. **Conclusions.** In this paper, a novel semi-supervised facial expression recognition algorithm on the condition of multi-pose was proposed. Our algorithm has brought transfer learning into semi-supervised learning, and solved the problem of multi-pose facial expression recognition efficiently. Simulation experimental results indicate that, the best recognition rates of the proposed algorithm are 82.68% and 87.71% on the RaFD database and BHU database, respectively. Therefore the proposed algorithm is competent for semi-supervised facial expression recognition on the condition of multi-pose.

Acknowledgment. The work described in this paper was supported by the National Natural Science Foundation of China under Grant No.30970780 and Ph.D. Programs Foundation of Ministry of Education of China under Grant No.20091103110005. The authors thank A.DEMRZ for providing the Splus code of the ASSEMBLE algorithm [16] used in our experiments, and would also like to thank Wei Wei for helping us to improve the linguistic quality of this paper.

REFERENCES

[1] S. W. Luo, *The Perception Computing of Visual Information*, Science Press, Beijing, China, 2010.

[2] I. Cohen, F. G. Cozman, N. Sebe, M. C. Cirelo, and T. S. Huang, Semisupervised learning of classifiers: theory, algorithms, and their application to human-computer interaction, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 12, pp. 1553-1567, 2004.

[3] M. F. A. Hady, M. Schels, F. Schwenker, and G. Palm, Semi-supervised facial expressions annotation using co-training with fast probabilistic tri-class SVMs, *Proc. of the 20th International Conference on Artificial Neural Networks*, pp. 70-75, 2010.

[4] K. Chen, and S. H. Wang, Semi-supervised learning via regularized boosting working on multiple semi-supervised assumptions, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 129-143, 2011.

[5] J. B. Li, S. C. Chu, J. S. Pan, and L. C. Jain, Multiple viewpoints based overview for face recognition, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, no. 4, pp. 352-369, 2012.

[6] W. Q Li, and D. S Chen, Multi-pose face recognition combining tensor face and manifold learning, *Proc. of International Conference on Computer Science and Automation Engineering*, pp. 543-547, 2011.

[7] J. H. Wang, J. You, Q. Lin, and Y. Xu, Orthogonal discriminant vector for face recognition across pose, *Pattern Recognition*, vol.45, no. 12, pp. 4069-4079, 2012.

[8] A. N. Li, S. G. Shan, and W. Gao, Coupled bias-variance tradeoff for cross-pose face recognition, *IEEE Trans. Image Processing*, vol. 21, no. 1, pp. 305-315, 2012.

[9] P. H. Lee, G. S. Hsu, Y. W. Wang, and Y. P. Hung, Subject-specific and pose-oriented facial features for face recognition across poses, *IEEE Trans. Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 42, no. 5, pp. 1357-1368, 2012.

[10] S. J. Pan, and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.

[11] W. Y. Dai, Q. Yang, G. R. Xue, and Y. Yu, Boosting for transfer learning, *Proc. of the 24th International Conference on Machine Learning*, pp. 193-200, 2007.

[12] Y. Freund, and R. E. Shapire, A decision-theoretic generalization of on-line learning and an application to boosting, *Proc. of the 2nd European Conference on Computational Learning Theory*, pp. 23-37, 1995.

[13] Y. L. Xue, X. Mao, and F. Zhang, Beihang university facial expression database and multiple facial expression recognition, *Proc. of the 5th International Conference on Machine Learning and Cybernetics*, pp. 3282-3287, 2006.

[14] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, Presentation and validation of the radboud faces database, *Cognition & Emotion*, vol. 24, no. 8, pp. 1377-1388, 2010.

[15] X. J Zhu, and Z. Ghahramani. Learning from labeled and unlabeled data with label propagation, *Technical Report CMU-CALD-02-107*, Carnegie Mellon University, pp. 1-7, 2002.

[16] K. P. Bennett, A. Demiriz, and R. Maclin, Exploiting unlabeled data in ensemble methods, *Proc. of the 8th ACM International Conference on Knowledge Discovery and Data Mining*, pp. 289-296, 2002.

[17] K. Chen, and S. H. Wang, Regularized boost for semi-supervised learning, *Proc. of the 21st International Conference on Neural Information Processing Systems*, pp. 281-288, 2007.