Enhancing Expression Recognition in the Wild with Unlabeled Reference Data

Mengyi Liu[1,2], Shaoxin Li[1,2], Shiguang Shan[1], Xilin Chen[1]

[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
[2]Graduate University of Chinese Academy of Sciences, Beijing 100049, China
{mengyi.liu, shaoxin.li, shiguang.shan, xilin.chen}@vipl.ict.ac.cn;

Abstract. Facial expression recognition is an important task in human-computer interaction. Some methods work well on "lab-controlled" data. However, their performances degenerate dramatically on real-world data as expression covers large variations, including pose, illumination, occlusion, and even culture change. To deal with this problem, large scale data is definitely needed. On the other hand, collecting and labeling wild expression data can be difficult and time consuming. In this paper, aiming at robust expression recognition in wild which suffers from the mentioned problems, we propose a semi-supervised method to make use of the large scale unlabeled data in two steps: 1) We enrich reference manifolds using selected unlabeled data which are closed to certain kind of expression. The learned manifolds can help smooth the variation of original data and provide reliable metric to maintain semantic similarity of expression; 2) To elevate the original labeled set for enhanced training, we iteratively employ the semi-supervised clustering to assign labels for unlabeled data and add the most discriminant ones into the labeled set. Experiments on the latest wild expression database SFEW and GENKI show that the proposed method can effectively exploit unlabeled data to improve the performance on real-world expression recognition.

1 Introduction

Facial expressions play important roles in our daily communications. Recognizing these expressions automatically has therefore become an active topic of research [1, 2]. After decades of development, facial expression recognition under controlled laboratory conditions is well solved by various promising methods [3–6]. However, when they were applied to unconstrained real-world data (e.g. internet images or personal photo albums), the performance may degrade dramatically. For this situation, study on realistic data is now regarded as a prospective issue which receives more and more attention.

The major challenge of real-world data analysis is handling the intra-class variation (e.g. pose, illumination, image resolution), which results in significant appearance change. To cope with the problem, in face recognition field, many recent researches [7–10] resort to an extra reference dataset. They construct certain similarity relationship between the query and the images in reference set, then process the recognition utilizing the corresponding labels of reference data. Ideally, if the large reference library exhaustively covers the data variation under different imaging conditions, the "associated" images found in the library can help transfer the face settings and decompose the large variation among probes.

This framework seems also feasible to real-world expression recognition. However, it is hard to find an extra reference set which has unambiguous expression definition as well as diverse appearance variation. Currently most of the expression datasets have been generated under constrained lab environments. Such "posed" expression data lacking of complex appearance variation is not capable in depicting the "spontaneous" expression data recorded in real-world. On the other hand, the labeling process of wild expression data can be expensive and time consuming. While collecting, without labeling, a large scale realistic data vary in different imaging conditions is not as difficult. Inspired by the strategy proposed in [7–10], we exploit data from such unlabeled set based on similarity relationships for smoothing the variation.

However unlabeled data have the limitation that they cannot provide the supervised information of categories to help recognition. Some related work, Cohen et al [11, 12] and Abdel et al [13] have proposed semi-supervised learning to effectively utilize the unlabeled data. In [12], the authors introduced a classification driven search algorithm for learning the best structure of the classifier. The unlabeled data in their work were collected from human posed expression. More recently, [13] applied co-training strategy using Tri-class SVM to predict unlabeled data that would be selected to enlarge the co-training set. However these two methods still coped with constrained expression recognition, and their unlabeled data are neither spontaneous nor in the wild. At this point, our problem is much more challenging and practical compared to these former works. In this paper, aiming at robust expression recognition in the wild, which suffers

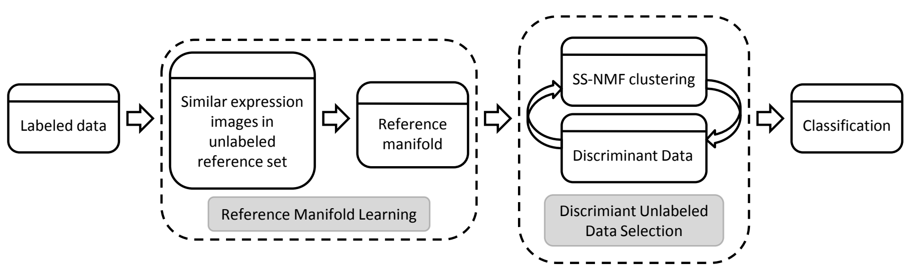


Fig. 1. The pipeline of the proposed method.

from the large intra-class variation and lack of real-world training data, we resort to a large scale unlabeled data to enhance the recognition in two steps: 1)

We construct reference manifolds using selected unlabeled data which are likely to be under certain kind of expression. The learned manifolds can help smooth the variation of original data and provide reliable metric to maintain semantic similarity of expression; 2) To augment the original labeled set for enhanced training, we iteratively employ the semi-supervised clustering to assign labels for unlabeled data, and add the most discriminant ones in the updated labeled set to boost the performance of our classification model. The whole procedure of our method is demonstrated in Fig. 1.

2 Reference Manifold Learning

In this section, we propose to extract similar expression samples from unlabeled set for reference manifold learning. Generally, one N-pixel face image can be considered as a point in the N-dimensional image space, and the variations of face images can be represented as low-dimensional manifolds embedded in the original image space [14]. Since the facial expression changes continuously in real-world, it is reasonable to assume that all expression images of an individual make a smooth manifold in the space [15].

Intuitively, face images with similar expressions should be lying in the local neighborhood on the expression manifold. However, in the original image space, due to significant variation caused by different poses, illumination and identities in real-world, the different expression images under the same condition may be "nearer" than the same expression images under different conditions (see Fig. 2, on the left, the red arrow stands for the distance between intra-class samples while the yellow arrow stands for the distance between inter-class samples). Obviously such data cannot help to learn the intrinsic structure of expression manifold. Following the idea in [8], we propose to select a certain number of most similar faces from the reference set to smooth the variation in original space. Given a reference set $R = \{r_1, r_2, ..., r_M\}$ and labeled expression data

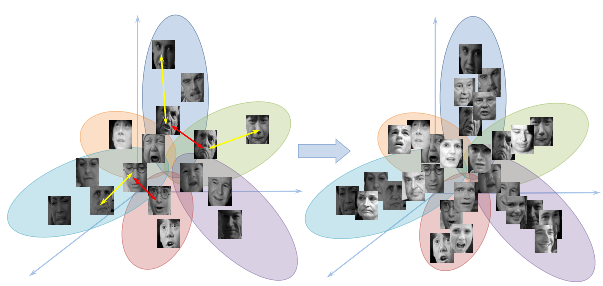


Fig. 2. The original "sparse" data have large intra-class scatter. After embedding in the reference manifold, each labeled data is surrounded by a number of unlabeled neighbors which can smooth the variation.

of c categories, represented as $E^1, E^2, ..., E^c$. We attempt to find similar faces in R according to the images of each expression category respectively. Suppose $E^t = \{e_1^t, e_2^t, ..., e_{n_t}^t\} (t \in \{1, 2, ..., c\})$, where n_t is the number of expression data in E^t. S_i^t is a set consist of the k nearest neighbors of e_i^t $(i \in \{1, 2, ..., n_t\})$ found in R. Based on the intuition that the more a reference image r_j is selected as nearest neighbor of images in E^t, the higher probability r_j has the t-th expression, we select expression data from these sets based on a frequency score,

$$F_j^t = \sum_{i=1}^{n_t} I(r_j \in S_i^t), \tag{1}$$

where $I(x)$ is an indicator function that $I(x) = 1$ if x is ture, and 0 otherwise. Then for each category E^t, we obtain a ranked list of r_j ordered by its frequency score F_j^t decreasingly. Only the first p^t images in the list are considered as the similar expression set of E^t, denoted as Ω^t. Finally, the whole selected set can be represented as $\Omega = \bigcup_{t=1}^{c} \Omega^t$, from which our reference manifold can be learned. Since the manifold can help smooth the variation and provide reliable metric to keep the semantic similarity of expression, the samples with the same expression are more likely to be grouped together after being projected on the manifold, as it demonstrates in Fig. 1.

For facial expression recognition, LPP/SLPP were successfully applied by [16, 17] to achieve better performance. Locality Preserving Projections (LPP)[18] shares similar data topology properties with nonlinear techniques (e.g. Isomap [19], LLE [14], and LE [20]). More crucially, LPP is defined throughout the space rather than just on training data, so that it has significant advantage in locating and explaining new data [16]. In this paper, we also explore LPP to learn the structure of the expression manifold. Unlike the works in [16, 17], we have separately learned multiple manifolds for global face and each local patch. Since the variations of local patches are relatively smaller compared to global image. It is more likely to find the most similar neighbors using patch-based searching. Additionally, the patch-based operation makes the added data of each patch coming from various samples in reference set, which further emphasize the information complementarity. Thus we can obtain much "denser" samples in the image space to construct smooth manifold. Fig. 3 shows the procedure of multiple reference manifold construction.

3 Discriminant Unlabeled Data Enhanced Training

Due to the diversity and large variation of real-world expression, training a robust classifier requires much larger amount of labeled data. As manually labeling wild expression data can be difficult and time consuming, we propose to augment the training set using unlabeled data with automatically assigned labels. However, currently available labeled samples are not sufficient for learning reliable classification model. Applying the model only trained on labeled set to assign labels for unlabeled data, the results may be severely biased to labeled set, thus cannot provide valuable information to enhance the original model.

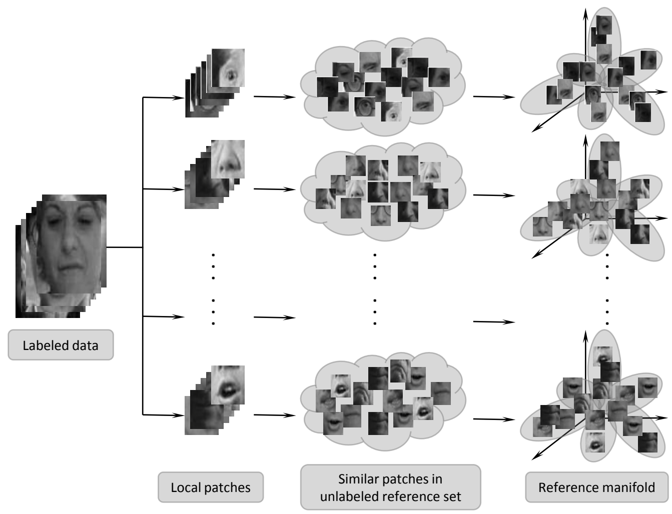


Fig. 3. The procedure of multiple reference manifolds construction (The figure only display the local manifold constructed with overlapping patches).

More crucially, since the distributions of both labeled and unlabeled data may not completely match each other, such bias model applied on selecting unlabeled data may cause performance degrading as increasing unlabeled data are added to training set. In this consideration, We employ semi-supervised clustering to assign labels for unlabeled data. Besides original training data's supervision, this framework also make use of large scale unlabeled data to obtain more accurate estimation of data distribution.

In this section, we first introduce the semi-supervised clustering algorithm SS-NMF [21], which is employed for our method. Then we formulate our algorithm of selecting discriminant unlabeled data according to their cluster label assigned by SS-NMF. Finally the enhanced training process is presented to explain how the unlabeled data work.

3.1 Semi-supervised clustering

There have been prior efforts on using provided class information to improve clustering [22]. In this paper, we applied Semi-Supervised Non-negative Matrix Factorization (SS-NMF) for its promising clustering performance as well as the low time complexity. Essentially, NMF can model widely varying data distributions and accomplish both hard and soft clustering simultaneously [21].

In our method, we cluster labeled and unlabeled data jointly using "must-link" and "cannot-link" constraints provided by the given class label. Suppose the labeled and unlabeled data consists of m and n objects respectively, with $k-$ *dimension* features extracted from each one (after manifold reduction introduced in section 2). Correspondingly, the data can be represented as two matrix $X_L \in$

$R^{m\times k}$ and $X_U \in R^{n\times k}$. For the whole data matrix $X = [X_L; X_U]$, it performs symmetric non-negative tri-factorization of the similarity matrix $A = XX^T \in R^{(m+n)\times(m+n)}$ as,

$$A \approx GSG^T, \tag{2}$$

where $G \in R^{(m+n)\times c}$ is the cluster assignment matrix, $S \in R^{c\times c}$ is the cluster centroid matrix that gives a compact $c \times c$ representation of X, and c is the number of clusters equals to the original expression class number.

Supervision is provided as two sets of pairwise constraints: must-link constraints C_{ML} and cannot-link constraints C_{CL}, which are accompanied by association violation cost matrix W. Here we use the class labels to construct C_{ML} and C_{CL}. For each pair (x_L^i, x_L^j) in X_L,

$$\begin{aligned}(x_L^i, x_L^i) \in C_{ML},\ s.t.\ y_L^i = y_L^j,\\ (x_L^i, x_L^i) \in C_{CL},\ s.t.\ y_L^i \neq y_L^j.\end{aligned} \tag{3}$$

The corresponding association violation cost matrix W_{reward} and $W_{penalty}$ are defined as,

$$\begin{aligned}W_{reward} = \{w_{ij} = A_{ij} - max\{A_{ij}\} | (x_L^i, x_L^j) \in C_{ML},\ s.t.\ g_L^i = g_L^j\},\\ W_{penalty} = \{w_{ij} = min\{A_{ij}\} - A_{ij} | (x_L^i, x_L^j) \in C_{CL},\ s.t.\ g_L^i = g_L^j\},\end{aligned} \tag{4}$$

where y_L^i is the class label of labeled data x_L^i, g_L^i is the cluster label of x_L^i, and w_{ij} is the penalty cost for violating the constraint between x_L^i and x_L^j. The objective function of SS-NMF is as follows:

$$J_{SS-NMF} = \min_{S\geqslant 0,\ G\geqslant 0} \parallel (A - W_{reward} + W_{penalty}) - GSG^T \parallel^2 . \tag{5}$$

An iterative procedure was proposed for the minimization by updating one factor S_{ih}(or G_{ih}) while fixing the others. Then we obtain the clustering results G for the further steps.

3.2 Discriminant unlabeled data selection

After SS-NMF clustering, we coarsely group the labeled and unlabeled data based on the structures of reference manifolds. However not all of these unlabeled data can be added for training, the reason is that besides the misclassified data, some data with correct cluster labels may be outliers or lie on boundary, thus the learned hyperplane may be bias due to the trade-off between maximizing the margin and minimizing the errors. For this situation, we propose to only select discriminant data to augment the training set according to the cluster labels assigned by SS-NMF. Here we define the "discriminant" data as which can enlarge the within-class similarity and reduce the between-class similarity of the original training data. Such data can adjust the original data distribution to help find the optimal hyperplane for classification.

We continue to use the representations that $X_L \in R^{m\times k}$ and $X_U \in R^{n\times k}$ denote the labeled and unlabeled data respectively. $L \in R^{m\times c}$ is the class assignment matrix of labeled data. By exploring SS-NMF clustering, we obtain the cluster assignment matrix of unlabeled data $G_U \in R^{n\times c}$.

As initialization, we calculate the mean within-class similarity $mean_w$ and mean between-class similarity $mean_b$ (Here we first compute the mean similarity (within-class or between-class) for each sample, then sum these similarities and normalize by the number of samples, i.e. m).

$$mean_w = \frac{1}{m} tr(\lambda_w X_L X_L^T (LL^T)^T), \tag{6}$$

$$mean_b = \frac{1}{m} tr(\lambda_b X_L X_L^T (H_L - LL^T)^T), \tag{7}$$

where $H_L = \mathbf{1}^{m\times m}$, and λ_w, λ_b are normalization matrices for averaging the similarities. Suppose k_i $(i \in \{1, 2, ..., c\})$ is the number of labeled samples in each class, i.e. $\sum_{i=1}^{c} k_i = m$. Then we have:

$$\lambda_w = diag\{\underbrace{k_1^{-1}, k_1^{-1}, ..., k_1^{-1}}_{k_1}, ..., \underbrace{k_c^{-1}, k_c^{-1}, ..., k_c^{-1}}_{k_c}\}, \tag{8}$$

$$\lambda_b = diag\{\underbrace{(m-k_1)^{-1}, ..., (m-k_1)^{-1}}_{k_1}, ..., \underbrace{(m-k_c)^{-1}, ..., (m-k_c)^{-1}}_{k_c}\}. \tag{9}$$

According to the clustering results, we can also calculate the within-class and between-class similarity matrices between each unlabeled data and the whole labeled set, denoted as Sim_w and Sim_b,

$$Sim_w = \lambda_w^* X_U X_L^T (G_U L^T)^T, \tag{10}$$

$$Sim_b = \lambda_b^* X_U X_L^T (H_U - G_U L^T)^T, \tag{11}$$

where $H_U = \mathbf{1}^{n\times m}$, and λ_w^*, λ_b^* are normalization matrices which have the similar forms as λ_w, λ_b respectively. Then the discriminant unlabeled data can be added in based on the rules that its within-class similarity is larger than $\boldsymbol{mean_w}$ while its between-class smaller than $\boldsymbol{mean_b}$. Please note that, after each iteration, the added unlabeled data will be treated as "labeled" ones, then new constraints will be constructed and the similarity matrix $\tilde{A}$ in SS-NMF, $\boldsymbol{mean_w}$, $\boldsymbol{mean_b}$ will be updated at the same time. In this way we can perform the selection process iteratively until these variables unchanged. The detailed selecting procedure are summarized in Algorithm. 1.

Finally, we have obtained the selected discriminant unlabeled data with their cluster labels to augment even refine the training set. In the enhanced training, global classifier and local classifiers are trained separately using SVM with original labeled data and each discriminant unlabeled data jointly. As it is mentioned in section. 1, the local training strategy makes the added data of each set coming from various samples in reference set, which further emphasizes the information complementarity. At last we apply weighted sum rule combining these sub-classifiers into one ensemble classifier to further improve the performance.

Algorithm 1 : Discriminant Unlabeled Data Selection Using SS-NMF

Input

Expression feature matrix of labeled data $X_L = [x_L^1; x_L^2; ...; x_L^m] \in R^{m\times k}$;
Expression feature matrix of unlabeled data $X_U = [x_U^1; x_U^2; ...; x_U^n] \in R^{n\times k}$;
Class assignment matrix of labeled data $L \in R^{m\times c}$;

Output

Selected discriminant unlabeled data $X_{US} = [x_U^{s_1}; x_U^{s_2}; ...; x_U^{s_d}] \in R^{d\times k}$;
The corresponding cluster assignment matrix $G_{US} = [g_U^{s_1}; g_U^{s_2}; ...; g_U^{s_d}] \in R^{d\times c}$;

Algorithm

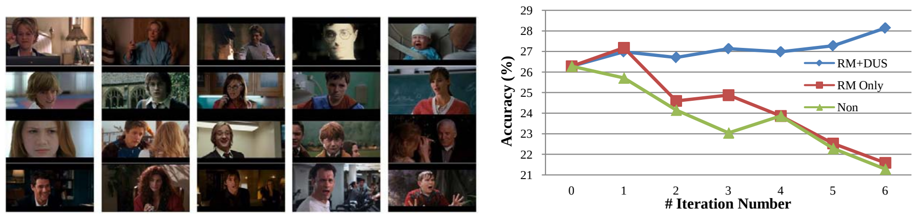
1. Initialize similarity matrix $A = XX^T$, where $X = [X_L; X_U]$;
2. Using the labeled data to construct $W_{penalty}$ and W_{reward}, then $\tilde{A} = A - W_{reward} + W_{penalty}$;
3. Obtain the cluster assignment matrix G:
   $G \leftarrow J_{SS-NMF}(\tilde{A} - GSG^T)$ $(G = [G_L; G_U])$,
   where G_U is cluster assignment matrix of unlabeled data.
4. Calculate mean similarities $\boldsymbol{mean_w}$ and $\boldsymbol{mean_b}$.
5. Calculate similarity matrices $\boldsymbol{Sim_w}$ and $\boldsymbol{Sim_b}$.
6. **if** $(Sim_w(i,i) > mean_w)\&\&(Sim_b(i,i) < mean_b)$
      Adding **discriminant unlabeled data** $\{x_U^i, g_{US}^i\}$ to labeled set.
   end if
7. **Update** $\tilde{A}$, and **repeat** steps 1-7 **until** $\tilde{A}$ unchanged.

4 Experiments

4.1 Experiments on SFEW

The Static Facial Expression in the Wild (SFEW) [23] which has been extracted from movies (see Fig. 4(a)) is the first attempt to build database depicting real-world or simulated real-world conditions for expression analysis. According to Strictly Person Independent (SPI) Protocol [23] for SFEW, the database is divided into two sets and the experiment is set to be two-fold. Each set contains seven subfolders corresponding to the seven expression categories. The sets were created in strict person independent manner that there is no overlap between training and testing set. The evaluation metrics for measuring performance are $accuracy$, $precision$, $recall\ and\ specificity$ (see definitions in [23]).

In the SPI baseline, the faces are localized using the Viola-Jones [24] face detector. Here we use the same way for fairly comparison. For classification, we use SVM [25] with RBF kernel as same as that in [23]. To demonstrate the effectiveness of our method, we simply apply HOG features [26] on the cropped face (size: 80x64). The unlabeled reference set in our experiments is LFW [27], which contains 13233 samples recorded in real-world conditions. χ^2 distance is employed for similarity computing during the searching in reference set. In the experiments, we first demonstrate the necessity of our two strategies (reference manifold learning in Sec. 2 and discriminant data selection in Sec. 3) in the whole procedure. In Fig. 4(b) the proposed method is denoted as "RM+DUS" (Reference Manifold + Discriminant Unlabeled data Selection); The result using all coarse-clustered unlabeled data based on reference manifold, without discriminant data selected, is denoted as "RM Only"; And "Non" represents the approach applying neither of the two strategies. To illustrate the relationship between recognition accuracy and the number of the added data, we iteratively augment the training set with unlabeled data and train new classifier combining the added data. Fig. 4(b) shows the accuracy of three methods at each iteration respectively.



(a) (b)

Fig. 4. (a) Sample images from the SFEW database [23]. (b) The "RM+DUS" shows an overall upward trend over iteration, while the results of "RM Only" and "Non" show that the performance degrade as increasing unlabeled data are added to training set.

To further improve the performance, we calculate the same features (HOG) on 12 overlapping patches (size: 32x32; sampling stride: 16 pixels) for local face analysis, as mentioned in Sec. 2. Besides the global-based results, Table. 1(a) also gives improved results combining global and local analysis. As it demonstrates, our method outperforms the baseline 19% [23] significantly. However, due to the tough imaging conditions, the faces obtained by automatic detector suffer from severe misalignment, so that the local patch-based strategy cannot work well. In this consideration, we perform image alignment using manually labeled 5 facial landmarks as the preprocessing. Based on these aligned data, we have achieved more significant improvement by using unlabeled data (see Table. 1(b)). Based

Table 1. The experimental results on SFEW (a) based on Viola Jones face detector, (b) based on 5 landmarks alignment result.

	Labeled data		Labeled + Unlabeled data	
	Global	Global + Local	Global	Global + Local
Fold1	25.71%	26.84%	27.97%	28.25%
Fold2	26.87%	27.75%	28.32%	30.35%
Avg	**26.29%**	**27.30%**	**28.14%**	**29.30%**

(a)

	Labeled data		Labeled + Unlabeled data	
	Global	Global + Local	Global	Global + Local
Fold1	30.23%	30.79%	31.64%	35.03%
Fold2	30.06%	30.06%	30.93%	34.68%
Avg	**30.14%**	**30.83%**	**31.29%**	**34.86%**

(b)

on the SPI protocol we compute the *precision*, *recall and specificity* scores

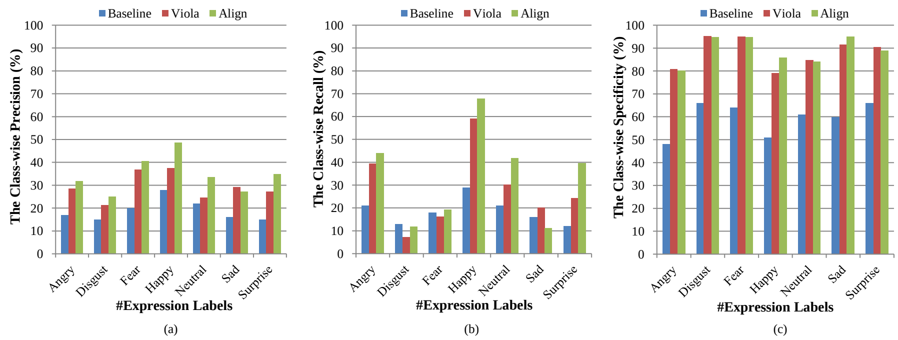


Fig. 5. The three class-wise evaluation metrics of different approaches. (a) precision, (b) recall, and (c) specificity.

respectively. Fig. 5 shows the comparison of the baseline in [23] and our method (where "Viola" stands for the results based on Viola Jones face detector; "Align" stands for results based on 5 landmarks alignment). Although the majority of the scores outperform the baseline significantly, there are still few cannot reach our expectation (the recall score of disgust, fear). The possible reason can be the lacking of typical expression samples in the reference set. In this case, to deal with some specific category, like disgust or fear, the selected unlabeled data may not be as discriminant as that in other categories. Future work should focus on this problem to improve the performance balanced to each class.

4.2 Experiments on GENKI

For real-life smile expression analysis, we perform our experiments on GENKI-4K [28], a subset of the images using in [29]. The database consists of 4000 web images captured in real-world (2162 "smile faces" and 1838 "non-smile faces"). Following the experimental settings in [30], the faces are normalized to 48x48 pixels based on eyes locations. Then we construct 4 equal subsets by sampling at intervals of 4 to adopt 4-fold cross-validation. We still apply HOG for feature descriptor and LFW for unlabeled reference dataset. Using linear kernel SVM (LIBSVM [25]) as in [30], we obtain results shown in Table. 2(a).

We observe only slight improvements by adding unlabeled data. The reason may be that our method is not suitable for this experimental condition with large size labeled training data. To make our experiments more challenging, we pick up all the 689 faces with large head pose ranging in $(-50°, -20°)$ and $(+20°, +50°)$ [30] to verify the proposed method. This subset satisfy our assumption that the data are relatively "sparse" with large variation. To adopt cross-validation, simply we construct 2 approximately equal subsets by sampling at intervals of 2 (Fold1: 345 for training and 344 for testing; Fold2: 344 for training and 345 for testing). The same feature descriptor and same classifier are explored in the experiments. As seen from Table. 2(b), which demonstrates that more significant improvement is achieved by using unlabeled data.

Table 2. The experimental results on GENKI. (a) The results of different approaches compared to [30] (on 4000 images). (b) The results based on only 689 large pose faces.

Approach				Accuracy(%)
Feature		Dimension	Classifier	
[31]	Gabor	23040	SVM	89.68±0.62
	LBP	944	SVM	87.10±0.76
	Raw Pixel Values	2304	SVM	81.45±0.32
	Pixel Comparison	500	Adaboost	89.70±0.45
HOG(labeled)		**1200**	**SVM**	**91.85±0.97**
HOG(labeled+unlabeled)		**1200**	**SVM**	**92.26±0.81**

(a)

	Labeled data	Labeled + Unlabeled data
Fold1	85.75%	88.37%
Fold2	87.82%	90.12%
Avg	**86.78±1.46%**	**89.24±1.23%**

(b)

5 Conclusion

Recognizing facial expression in the wild is an interesting and challenging problem in many applications. To cope with such complicated problem, one of the solutions may be training on a large amount of labeled data, which are difficult to collect in real-world scenarios. In this paper, we propose to augment small size labeled set with unlabeled data by iteratively searching the “discriminant” samples under a semi-supervised framework. To evaluate the method, we perform our experiments on the latest wild expression databases SFEW and GENKI. The results show that the proposed method can effectively exploit unlabeled data to improve the performance on real-world expression recognition.

Acknowledgement. This work is partially supported by Natural Science Foundation of China (NSFC) under contract Nos. U0835005, and 60832004; and Beijing Natural Science Foundation (New Technologies and Methods in Intelligent Video Surveillance for Public Security) under contract No. 4111003.

References

1. Pantic, M., Rothkrantz, L.: Automatic analysis of facial expressions: The state of the art. IEEE T PAMI **22** (2000) 1424–1445
2. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE T PAMI **31** (2009) 39–58
3. Shan, C., Gong, S., McOwan, P.: Facial expression recognition based on local binary patterns: A comprehensive study. IVC **27** (2009) 803–816
4. Dhall, A., Asthana, A., Goecke, R., Gedeon, T.: Emotion recognition using phog and lpq features. In: FG. (2011) 878–883
5. Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., Metaxas, D.: Learning active facial patches for expression analysis. In: CVPR. (2012) 2562–2569
6. Chew, S., Lucey, S., Lucey, P., Sridharan, S.: Improved facial expression recognition via uni-hyperplane classification. In: CVPR. (2012) 2554–2561
7. Kumar, N., Berg, A., Belhumeur, P., Nayar, S.: Attribute and simile classifiers for face verification. In: ICCV. (2009) 365–372

8. Yin, Q., Tang, X., Sun, J.: An associate-predict model for face recognition. In: CVPR. (2011) 497–504
9. Schroff, F., Treibitz, T., Kriegman, D., Belongie, S.: Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison. In: ICCV. (2011) 2494–2501
10. Su, Y., Shan, S., Chen, X., Gao, W.: Adaptive generic learning for face recognition from a single sample per person. In: CVPR. (2010) 2699–2706
11. Cohen, I., Sebe, N., Gozman, F., Cirelo, M., Huang, T.: Learning bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In: CVPR. Volume 1. (2003) 595–601
12. Cohen, I., Sebe, N., Cozman, F., Huang, T.: Semi-supervised learning for facial expression recognition. In: ACM SIGMM Workshop on MIR. (2003) 17–22
13. Abdel-hady, M., Schels, M., Schwenker, F., Palm, G.: Semi-supervised facial expressions annotation using co-training with fast probabilistic tri-class svms. ANN (2010) 70–75
14. Saul, L., Roweis, S.: Think globally, fit locally: unsupervised learning of low dimensional manifolds. JMLR **4** (2003) 119–155
15. Chang, Y., Hu, C., Turk, M.: Manifold of facial expression. In: AMFG. (2003) 28–35
16. Shan, C., Gong, S., McOwan, P.: Appearance manifold of facial expression. CVHCI (2005) 221–230
17. Xiao, R., Zhao, Q., Zhang, D., Shi, P.: Facial expression recognition on multiple manifolds. PR **44** (2011) 107–116
18. He, X., Niyogi, P.: Locality preserving projections. NIPS **16** (2004) 153–160
19. Tenenbaum, J., De Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science **290** (2000) 2319–2323
20. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. NIPS **14** (2001) 585–591
21. Chen, Y., Rege, M., Dong, M., Hua, J.: Non-negative matrix factorization for semi-supervised data clustering. KIS **17** (2008) 355–379
22. Kulis, B., Basu, S., Dhillon, I., Mooney, R.: Semi-supervised graph clustering: a kernel approach. In: ICML. (2005) 457–464
23. Dhall, A., Goecke, R., Lucey, S., Gedeon, T.: Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In: ICCV Workshop on BeFIT. (2011) 2106–2112
24. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR. Volume 1. (2001) 511–518
25. Chang, C., Lin, C.: Libsvm: a library for support vector machines. ACM TIST **2** (2011) 27
26. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR. Volume 1. (2005) 886–893
27. Huang, G., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. Technical Report07-49, University of Massachusetts, Amherst (2008)
28. GENKI-4K: (http://mplab.ucsd.edu, the mplab genki dataset: Genki-4k subset)
29. Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M., Movellan, J.: Toward practical smile detection. IEEE T PAMI **31** (2009) 2106–2111
30. Shan, C.: Smile detection by boosting pixel differences. IEEE T IP **21** (2012) 431–436