Weakly Supervised Metric Learning towards Signer Adaptation for Sign Language Recognition

Fang Yin[1,2]
fang.yin@vipl.ict.ac.cn

Xiujuan Chai[1,2]
xiujuan.chai@vipl.ict.ac.cn

Yu Zhou[3]
zhouyu@iie.ac.cn

Xilin Chen[1,2]
xlchen@ict.ac.cn

[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS
Beijing, 100190, China

[2] Cooperative Medianet Innovation Center, China

[3] Institute of Information Engineering, CAS
Beijing, 100093, China

Abstract

Inter-signer variation is one of the challenging factors in real application of Sign Language Recognition(SLR). To tackle this problem, this paper proposes a weakly supervised metric learning framework to realize signer adaptation with some unlabeled sign data of the new signer. Concretely speaking, through the labeled data of several different signers, a generic distance metric can be learnt. Then the key step is to adapt the generic metric to the new signer according to the collected unlabeled samples. Clustering constraint and manifold constraint are considered together to realize the weakly supervised metric learning. In our implementation, a novel fragment-based feature is designed to describe each sign by fusing both trajectory and hand shape features, which is also proved more discriminative than the frame-based multi-modal feature. Extensive experiments on our collected large vocabulary datasets convincingly show that the proposed method is effective for signer adaptation and outperforms the state-of-the-art methods on signer-independent SLR.

1 Introduction

Automatic Sign Language Recognition (SLR), as one of the important problems in pattern recognition field, has attracted more and more researchers' attention and achieved rapid progress in over twenty years. SLR systems usually work well if the training and probe data come from the same signer. But when confronted with a new signer, the performance will decrease dramatically because of the big variations among different signers. Therefore, the signer-independent SLR has become one of the bottleneck problems which blocks the recognition algorithms to be practical. Although the user of the SLR system is usually unknown, fortunately, more and more data of this new signer can be collected sustainably.

These unlabeled data can help to refine the training model and thus improve the recognition performance. As a result, the SLR system will become more and more easy-to-use for this signer because of the adaptation. This paper just focuses on the signer adaptation and tries to tackle the signer-independent SLR problem.

In recent years, many attempts have been made to cope with the inter-signer variation through signer adaptation. Borrowing from speaker adaptation in speech recognition, eigen-voice(EV) [8], maximum likelihood linear regression (MLLR) [9] and maximum a posteriori (MAP) [6] are used in SLR signer adaptation. Ong et al.[14, 15] firstly used MAP in sign-er adaptation for gesture recognition. Von et al [16] combined MLLR and MAP to realize signer adaptation for isolated SLR. Later they extended their work to continues SLR and combined EV, MLLR and MAP [17]. Farhadi *et al.* [5] used the adaptation data from the new signer to generate a comparative feature. Oya and Lale[1] presented a multi-class clas-sification strategy for Fisher scores and applied it to sign language recognition. They also proposed a signer adaptation method during the score space selection phase. The above mentioned methods are all supervised adaptation, which means that the adaptation data from the new signer must be labeled. But in realistic practice, collecting labeled data is laborious and time-consuming. Comparatively, unsupervised adaptation is more natural and friendly for the easy collection of plenty of unlabeled data from the daily use. However, the work in this category is relatively rare. Zhou et al [23] proposed an unsupervised adaptation method based on hypothesis comparison guided cross validation.

Most of the previous work used HMM and its variations. However, their performances were not satisfied especially confronted with insufficient training samples. Recently, many novel techniques were introduced into SLR area, such as sequential pattern trees[13], trajec-tory matching[3] and CRF[7], etc. It is well known metric learning is an important topic in machine learning and widely used in many applications[21]. Studies have proved that learn-ing a proper distance metric could help to improve the performance of classification[20]. Therefore, this paper proposes a new unsupervised signer adaptation method based on dis-tance metric learning.

Metric learning can be categorized as supervised metric learning, unsupervised metric learning, semi-supervised metric learning and weakly supervised metric learning. In super-vised metric learning[4, 20], the labels of training data are known while in unsupervised metric learning[12], the labels of the training data are totally unknown. Semi-supervised metric learning[11] is a combination of supervised metric learning and unsupervised metric learning. In weakly supervised metric learning[10], the labels for training data are uncertain.

In this paper, we introduce metric learning into SLR for the first time and propose a signer adaption framework to address signer-independent SLR. For adapting the general model to the new signer, both clustering and manifold constraints are considered in the adaptive distance metric optimization.

The contribution of our work mainly lies in three-folds. Firstly, a Weakly Supervised Metric Learning(WSML) framework is proposed, which combines the clustering and mani-fold constraints simultaneously. Secondly, the general framework is applied on signer adap-tation and achieves good performance. Thirdly, a fragment based feature is designed for sign language representation and the effectiveness is verified in large vocabulary datasets.

The remainder of the paper is organized as follows. Section 2 is the formulation and optimization of our proposed weakly supervised metric learning method. Section 3 is its application on signer adaptation. Wide experiments are reported in Section 4 and finally is the conclusion in Section 5.

2 Weakly Supervised Metric Learning

Let $X_l = \{x_1, ..., x_l\}$ be the labeled training data and $Y = \{y_1, y_2, ..., y_l\}$ is the label set where y_i is the label of x_i. $X_u = \{x_{l+1}, ..., x_{l+u}\}$ represents the unlabeled adaptation data sampled from a different probability distribution. $X = [x_1, ..., x_l, x_{l+1}, ..., x_{l+u}]$ is the data matrix of all samples. The target is to learn a distance metric $D\colon X \times X \longrightarrow \mathbb{R}$ which is robust to both X_l and X_u. In the Mahalanobis distance, D can be determined by a positive semi-definite matrix M:

$$D(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j). \tag{1}$$

So the goal of metric learning is to learn a matrix M for distance metric D.

Our proposed Weakly Supervised Metric Learning(WSML) framework can be decomposed into two key steps. First is to learn a generic metric from the given labeled data. Then is to realize the distance metric adaptation by considering the clustering and manifold constraints with the unlabeled data.

2.1 Learning a Generic Distance Metric

To learn a generic distance metric, the labeled data are used under clustering assumption with classical large margin hinge loss. Specifically, the distances between data points within the same cluster(with same label) should be minimized and the distances between data points from different clusters(with different labels) should be maximized. Here we define the index set with same labels as $S_g = \{(i, j) | y_i = y_j, x_i, x_j \in X_l\}$ and the index triplet $B_g = \{(i, j, k) | y_i = y_j, y_i \neq y_k, x_i, x_j, x_k \in X_l\}$. The objective function is

$$\min J_g = \min_{M_g} \sum_{(i,j) \in S_g} D_g(x_i, x_j) + \sum_{(i,j,k) \in B_g} [1 + D_g(x_i, x_k) - D_g(x_i, x_j)]_+ \tag{2}$$

$$s.t. M_g \geq 0,$$

where the term $[z]_+ = max(z, 0)$ denotes the standard hinge loss and M_g is the metric matrix for distance metric D_g.

The objective function (2) can be rewritten by incorporating Eq.(1):

$$\begin{aligned}
\min J_g = \min_{M_g} &\sum_{(i,j) \in S_g} (x_i - x_j)^T M_g (x_i - x_j) \\
&+ \sum_{(i,j,k) \in B_g} [1 + (x_i - x_k)^T M_g (x_i - x_k) - (x_i - x_j)^T M_g (x_i - x_j)]_+ \\
= \min_{M_g} &\sum_{(i,j) \in S_g} tr(M_g C_{ij}) + \sum_{(i,j,k) \in B_g} [1 + tr(MC_{ik}) - tr(M_g C_{ij})]_+,
\end{aligned} \tag{3}$$

$$s.t. M_g \geq 0.$$

where $C_{ij} = (x_i - x_j)(x_i - x_j)^T$.

The expression (3) can be optimized by gradient descent algorithm. The detailed optimization algorithm is similar with [20].

2.2 Distance Metric Adaptation

The generic distance metric is valid for the data coming from the same distribution with the training data. However, for a different data distribution, the distance metric would be failed.

This section will give the details on our proposed weakly supervised metric learning method. In our distance metric adaptation, two constraints are considered in the optimization, which are clustering constraint and manifold constraint respectively.

2.2.1 Clustering Constraint

With the generic distance metric D_g determined by matrix M_g and the labeled data set X_l, a classifier is trained. Thus the labels of the unlabeled data can be predicted. We hope that the generic distance metric can be adaptive to the data from a different distribution through the predicted labels, although they are uncertain. For each unlabeled data x_i, the classifier gives a predicted label y_i and belief b_i, where $0 \leq b_i \leq 1$. Here the label y_i is not precise so the metric learning method with these uncertain labels is weakly supervised. The objective function of clustering constraint is

$$J_c = \sum_{(i,j) \in S_c} b_i D(x_i, x_j) + \sum_{(i,j,k) \in B_c} b_i [1 + D(x_i, x_k) - D(x_i, x_j)]_+ \qquad (4)$$

where S_c is a index set defined as $S_c = \{(i,j) | y_i = y_j, x_i \in X_u, x_j \in X_l\}$. While B_c is a triplet set $B_c = \{(i,j,k) | y_i = y_j, y_i \neq y_k, x_i \in X_u, x_j, x_k \in X_l\}$.

2.2.2 Manifold Constraint

Here, one point should be noted that only partial data of the unlabeled set can be assigned the correct labels with high believes. Thus we need to explore much more information provided by lots of unlabeled data for the effective adaptation. To further investigate the topological structure of the unlabeled data, manifold assumption is considered. Specifically, if two data points x_i and x_j are close in the intrinsic geometry of Euclidean distance, then they should also be close to each other in the new Mahalanobis distance. A straightforward expression for manifold assumption is

$$J_m = \sum_{i=l+1}^{u} \sum_{j=l+1}^{u} (x_i - x_j)^T M (x_i - x_j) W_{ij}, \qquad (5)$$

where W is a weight matrix. $W_{ij} = 1$ if x_i is among the $k-$nearest neighbors of x_j or x_j is among the $k-$nearest neighbors of x_i. Otherwise, $W_{ij} = 0$.

2.2.3 Formulation and Optimization

The final objective function of metric adaptation can be formulated by incorporating expressions (4) and (5) into (2).

$$\min J = \min(J_g + \alpha J_c + \beta J_m), \qquad (6)$$

where α and β are the weights corresponding to clustering constraint and manifold constraint respectively.

Define an active triplet set $B_{gs} \subset B_g$, such that $(i,j,k) \in B_{gs}$ could trigger the hinge loss of J_g in Equation (6). While $B_{cs} \subset B_c$ is an active triplet set for J_c in Equation (6). To solve
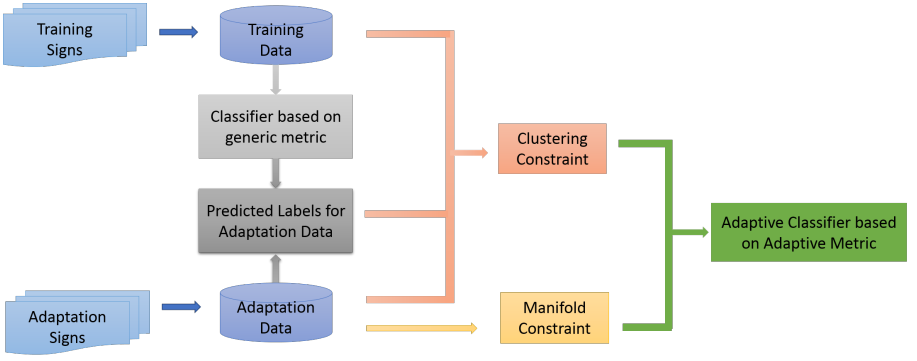
Figure 1: The pipeline of the signer adaptation.

the problem (6), optimization by gradient descent algorithm is also adopted. The gradient is

$$
G = \frac{\partial J}{\partial M} = \sum_{(i,j) \in S_g} C_{ij} + \sum_{(i,j,k) \in B_{gs}} (C_{ik} - C_{ij})
$$
$$
+ \sum_{(i,j) \in S_c} b_i C_{ij} + \sum_{(i,j,k) \in B_{cs}} b_i (C_{ik} - C_{ij}) + \sum_{i=l+1}^{u} \sum_{j=l+1}^{u} C_{ij} W_{ij}.
$$
(7)

The detailed optimization is shown in Algorithm 1. The parameter γ in step 6 is the gradient step-size and $\gamma > 0$.

Algorithm 1 Optimization of Problem (6).

1: $M_0 := I$ {Initialize with the identity matrix}
2: $t := 0$ {Initialize counter}
3: **while** not converged **do**
4:    update $B_{gs}^{(t+1)}$ and $B_{cs}^{(t+1)}$
5:    compute G_{t+1} based on Equation (7)
6:    $M_{t+1} := M_t - \gamma G_{t+1}$
7:    $t := t+1$
8: **end while**
9: **return** M_t

---

3 Application on Signer Adaptation

In this section we will introduce how to apply the weakly supervised metric learning on signer adaptation. The framework is described firstly. Then a fragment based feature is introduced. Finally the detailed implementation of training and recognition is presented.

3.1 Framework

The pipeline of the signer adaptation is illustrated by Figure 1. Concretely speaking, the features are extracted from the labeled signs and the unlabeled adaptation signs. With the

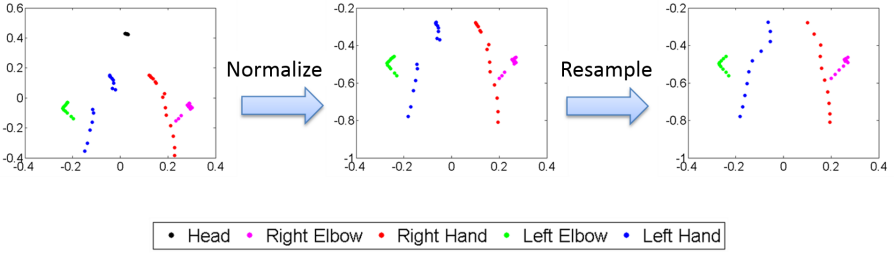• Head   • Right Elbow   • Right Hand   • Left Elbow   • Left Hand

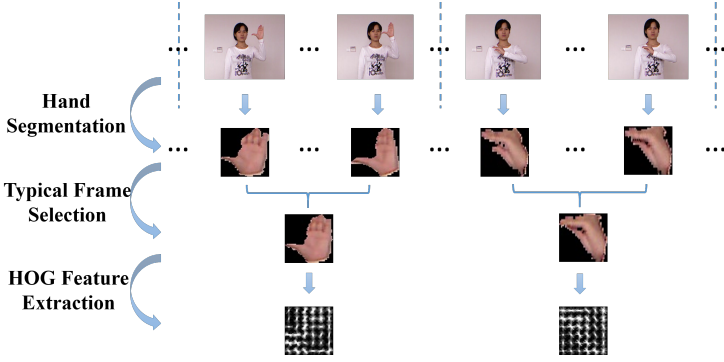Figure 2: Trajectory feature extraction.

Figure 3: Illustration for fragment-based feature extraction.

features of labeled training data, a generic distance metric can be learnt, with which a classifier is trained. Then the labels of unlabeled data are predicted with believes. In the adaptation stage, both the labeled and unlabeled training data are used for adaptive metric learning. Specifically, the labeled data and the unlabeled data with their predicted labels are used together for clustering constraint, while the unlabeled data are used for manifold constraint. The classifier based on the adaptive metric is suitable for the new signer and is used for recognition.

3.2 Feature Extraction

Intrinsically, sign language can be regarded as a temporal ensemble of some action subunits. Inspired from this point, a fragment based feature is proposed to characterize the approximate subunits. A sign is partitioned into N fragments equally. We extract trajectory and hand shape of each sign fragment Then they are fused into a whole feature vector which can describe both the movement feature and the hand shape feature of the sign fragment.

Figure 2 is the illustration of trajectory feature extraction. With the help of the depth data provided by Kinect, the 3D coordinates of 5 skeleton joints (left hand, right hand, left elbow, right elbow and head) are used for trajectory representation. The trajectory feature extraction are formed by normalization step and resampling step. In the normalization step, the head point is used as a reference(the origin). The points of each joint in the fragment could form a trajectory. Then resampling is performed along each trajectory by equidistant linear interpolation. So L equally distributed points can be gotten in each trajectory. In our

implementation, $L = 10$, and the dimension of the trajectory feature for each fragment is 120 ($4 \times 10 \times 3$).

Figure 3 is the illustration of hand shape feature extraction. The hands in each frame are segmented based on the skin color model and the depth constraint. HOG[3] feature is extracted from the segmented hand. Then PCA technique is adopted to reduce the dimension from 324 to 51. In each fragment, the frame which is most similar to the average of all the hand regions in the fragment is selected as the typical frame to describe this fragment. Here, we use Euclidean distance to define the distances between the HOG features.

3.3 Training and Recognition

In the training stage, all the feature vectors of labeled and unlabeled sign data are used as input. Then the adaptive metric M for new signer is learnt. Since M is a positive semi-definite matrix, it could be decomposed

$$M = L^T L. \tag{8}$$

So the Mahalanobis distance between two points(x_i and x_j) with matrix M is

$$
\begin{aligned}
D(x_i, x_j) &= \sqrt{(x_i - x_j)^T M (x_i - x_j)} \\
&= \sqrt{(x_i - x_j)^T L^T L (x_i - x_j)} = \sqrt{(Lx_i - Lx_j)^T (Lx_i - Lx_j)}.
\end{aligned}
\tag{9}
$$

Equation (9) shows that the Mahalanobis distance with matrix M_m is equivalent to an Euclidean distance of the data in the projected space transformed by matrix L. Therefore, all the training data are projected with L for the subsequent recognition.

In the recognition stage, the simple KNN classifier is adopted. For a probe data point x_p, it is also projected with L. Then the distance between x_p and any training data point $x_i \in X$ can be calculated by Equation (9). The 3 nearest neighbors from the training data of the probe data are selected and vote for the final recognition result.

4 Experiments

In order to evaluate the effectiveness of the proposed weakly supervised metric learning method, experiments on signer-independent SLR are conducted. The details are given below.

4.1 Performance Evaluation

4.1.1 Dataset

We collected the Sign Language dataset using Microsoft Kinect. When capturing, each signer stands before the Kinect at a distance around 1.5 meters to ensure his/her body is completed captured in the screen. Our dataset contains 1000 signs of standard Chinese Sign Language performed by 7 deaf students. Each signer performs each sign only once. Here we define $D_1 - D_7$ as group ID to indicate the 1000 signs performed by different signers.

Table 1: Evaluation on parameter α and β

Accuracy(%)	$\beta = 0$	$\beta = 5$	$\beta = 7$	$\beta = 10$	$\beta = 15$
$\alpha = 0$	70.0	70.2	70.4	70.5	69.7
$\alpha = 5$	70.2	70.4	70.6	71.0	69.8
$\alpha = 7$	70.4	70.8	70.8	71.2	69.8
$\alpha = 10$	71.1	71.5	71.8	72.3	70.8
$\alpha = 15$	70.2	70.2	70.5	70.5	69.4

Table 2: Experimental Results with different features

Group	D_1	D_2	D_3	D_4	D_5	D_6	D_7	Ave.
WSML(Frame)	65.4	70.2	64.1	62.6	65.4	75.5	63.0	**66.6**
WSML(Frag)	72.3	75.1	70.3	77.2	71.0	79.9	69.2	**73.5**

4.1.2   Evaluation on Different Parameters

In this section, we will first evaluate the performance of the proposed WSML with different parameters in Equation (6) on above-mentioned dataset.

In this experiment, we use the data of $D_2 - D_7$ as labeled training data and thus D_1 is the test signer. Half of the 1000 signs of the test signer is taken as the unlabeled adaptation data and the rest 500 samples are the test data.

We evaluate the performance with different values of parameters α and β. Table 1 gives the performance and it shows that when $\alpha = 10$ and $\beta = 10$, we achieve the highest accuracy. Thus in the following experiments, we fix the parameters to be $\alpha = 10$ and $\beta = 10$.

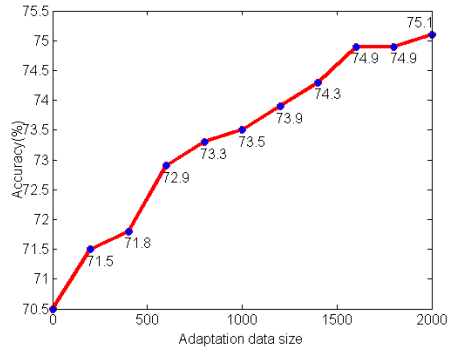### 4.1.3   Evaluation on Different Features

In this part, the target is to verify the effectiveness of the proposed fragment-based feature. In this experiment, the frame-based feature is taken as the comparison under the same WSML framework. In the implementation, skeleton pairwise feature[19] and hand shape's HOG feature are extracted from each frame. To maintain the definite dimension, a resampling technique is applied on the features and finally 15-frame features are generated by linear interpolation and the final dimension is 915. The dimension of our fragment-based feature is 855. The weakly supervised metric learning with our proposed fragment-based feature and the frame-based feature are referred as WSML(Frag) and WSML(Frame) respectively. In order to get more convincing results, the experiments here and in the following sections are all conducted by leave-one-out cross validation strategy. In each fold, half of the data from the adaptive signer is used for adaptation, and the remaining 500 samples are used for test. The adaptation data and the test data will exchange so that the final recognition rate is the average recognition rate over the whole 1000 samples of the adaptive signer. Table 2 shows the comparisons with two different features. We can obviously see that at least in our framework, the fragment-based feature is much more effective than the traditional frame-based feature with 6.9 percentage points performance improvement.

Table 3: Effectiveness Verification.

Group	D_1	D_2	D_3	D_4	D_5	D_6	D_7	Ave.
ML	70.0	72.1	69.0	75.6	69.1	77.7	66.2	**71.3**
ML-Clus	71.1	73.6	69.5	75.7	68.8	80.8	67.4	**72.4**
ML-Mani	70.5	73.7	69.0	76.6	67.7	79.6	69.3	**72.3**
WSML	72.3	75.1	70.3	77.2	71.0	79.9	69.2	**73.5**
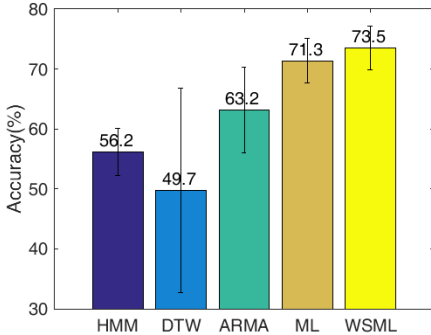
Figure 4: The accuracies of our proposed method and the baseline methods are given in (a). (b) shows the accuracies of our proposed method with different adaptation data sizes.

4.1.4 Evaluation on Different Constraints

In this section, we will show the effect of clustering constraint and manifold constraint respectively. The configurations are the same as above experiment. The results are given in Table 3, which includes the generic metric learning (ML), metric learning adaptation with clustering constraint (ML-Clus), metric learning adaptation with manifold constraint (ML-Mani), and the proposed weakly supervised metric learning with both constraints(WSML). From this table, it can be seen that the two different constraints are both helpful for improving the performance of the signer-independent SLR comparing with the generic ML method. When combining the two constraints, the signer adaption is more effective than using them separately.

4.2 Comparisons with Other Methods

In this part, we will compare our method with some classic and state-of-the-art methods. We set HMM[18] and Dynamic Time Warping (DTW) as our baseline methods because they are typical models for sign language recognition. The input of HMM and DTW is the dense frame-based features composed by concatenating hand shape feature and trajectory feature. Besides HMM and DTW, ARMA[22] is also used as a comparison for its good performance in handling the temporal sequential problems. Here, one point should be mentioned that in our experiments, we use the source codes from [18] and [22] to realize HMM and ARMA methods respectively and use the matlab source code to realize DTW based method. In

the methods mentioned above, the data from the new signer is directly fed into the recognition algorithm and get the result with no adaptation. All experimental results are shown in Figure 4(a), from which, we can see the distinctive advantage of our proposed method over others in such a large vocabulary dataset. The figure shows that comparing with classic methods (HMM and DTW), the accuracy is improved by 15-20 percentage points. And the accuracy of our WSML method is about 10 percentage points higher that the state-of-the-art method(ARMA). At the same time, the standard deviations of our methods are relatively small and it means that our method has a stable performance. Comparing with the generic metric learning(ML), the proposed WSML method could further increase the accuracy with more than 2 percentage points.

4.3 Effect with Different Sizes of Adaptation Data

The experimental results in Section 4.2 show that the improvement of WSML comparing with ML is not significant. One of the reasons is the limited adaptation data. In this section, we will show the accuracy trend with the increasing of adaptation data size. In this experiment, all data of $D_1 - D_7$ are used for training. Three groups of data are collected from another signer, who is regarded as the test signer. Among the three groups of data, one group (1000 samples) is used as test data and the others (2000 samples) are adaptation data. We gradually add different amount of unlabeled data into adaptation, and get the recognition performance for the same 1000 testing samples. Figure 4(b) shows the experimental results and it can be clearly seen that the performance is better and better with more and more adaptation data. The accuracy improvement is 4.6 percentage points with adaptation data size of 2000. It can be predicted that the performance of our system can be close to that of signer-dependent SLR case when enough sign data are collected for signer adaptation.

5 Conclusion

In this paper, a weakly supervised metric learning framework is proposed for signer adaptation. By considering the clustering constraint and manifold constraint simultaneously, the information from the unlabeled data are deeply investigated and the generic metric is adapted to the new data distribution. The performance of signer-independent SLR can be improved with the WSML framework applied to signer adaptation. Widely experiments on large vocabulary datasets convincingly show that the proposed method outperforms state-of-the-art methods.

6 Acknowledgement

References

[1] Oya Aran and Lale Akarun. A multi-class classification strategy for fisher scores: Application to signer independent sign language recognition. *Pattern Recognition*, 43 (5):1776–1788, 2010.

[2] Xiujuan Chai, Guang Li, Yushun Lin, Zhihao Xu, Yili Tang, Xilin Chen, and Ming Zhou. Sign language recognition and translation with kinect. In *IEEE Conf. on AFGR*, 2013.

[3] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[4] Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*, pages 209–216. ACM, 2007.

[5] Ali Farhadi, David Forsyth, and Ryan White. Transfer learning in sign language. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[6] Jean-Luc Gauvain and Chin-Hui Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *Speech and audio processing, ieee transactions on*, 2(2):291–298, 1994.

[7] WW Kong and Surendra Ranganath. Towards subject independent continuous sign language recognition: A segment and merge approach. *Pattern Recognition*, 47(3): 1294–1308, 2014.

[8] Roland Kuhn, Jean-Claude Junqua, Patrick Nguyen, and Nancy Niedzielski. Rapid speaker adaptation in eigenvoice space. *Speech and Audio Processing, IEEE Transactions on*, 8(6):695–707, 2000.

[9] Christopher J Leggetter and Philip C Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models. *Computer Speech & Language*, 9(2):171–185, 1995.

[10] Yadong Mu, Jialie Shen, and Shuicheng Yan. Weakly-supervised hashing in kernel space. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3344–3351. IEEE, 2010.

[11] Gang Niu, Bo Dai, Makoto Yamada, and Masashi Sugiyama. Information-theoretic semi-supervised metric learning via entropy regularization. *Neural computation*, 26 (8):1717–1762, 2014.

[12] X Niyogi. Locality preserving projections. In *Neural information processing systems*, volume 16, page 153. MIT, 2004.

[13] Eng-Jon Ong, Oscar Koller, Nicolas Pugeault, and Richard Bowden. Sign spotting using hierarchical sequential patterns with temporal intervals. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1931–1938. IEEE, 2014.

[14] Sylvie CW Ong and Surendra Ranganath. Deciphering gestures with layered meanings and signer adaptation. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 559–564. IEEE, 2004.

[15] Sylvie CW Ong, Surendra Ranganath, and Yedatore Venkatakrishnaiya Venkatesh. Understanding gestures with systematic variations in movement dynamics. *Pattern recognition*, 39(9):1633–1648, 2006.

[16] Ulrich Von Agris, Daniel Schneider, Jörg Zieren, and K-F Kraiss. Rapid signer adaptation for isolated sign language recognition. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 159–159. IEEE, 2006.

[17] Ulrich Von Agris, C Blomer, and K-F Kraiss. Rapid signer adaptation for continuous sign language recognition using a combined approach of eigenvoices, mllr, and map. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[18] Chunli Wang, Wen Gao, and Shiguang Shan. An approach based on phonemes to large vocabulary chinese sign language recognition. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 411–416. IEEE, 2002.

[19] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1290–1297. IEEE, 2012.

[20] Kilian Q Weinberger, John Blitzer, and Lawrence K Saul. Distance metric learning for large margin nearest neighbor classification. In *Advances in neural information processing systems*, pages 1473–1480, 2005.

[21] Eric P Xing, Michael I Jordan, Stuart Russell, and Andrew Y Ng. Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems*, pages 505–512, 2002.

[22] Chunyan Xu, Tianjiang Wang, Junbin Gao, Shougang Cao, Wenbing Tao, and Fang Liu. An ordered-patch-based image classification approach on the image grassmannian manifold. *Neural Networks and Learning Systems, IEEE Transactions on*, 25(4):728–737, 2014.

[23] Yu Zhou, Xiaokang Yang, Yongzheng Zhang, Xiang Xu, Yipeng Wang, Xiujuan Chai, and Weiyao Lin. Unsupervised adaptive sign language recognition based on hypothesis comparison guided cross validation and linguistic prior filtering. *Neurocomputing*, 149, Part C(0):1604 – 1612, 2015. ISSN 0925-2312.