

Cluster-Pairwise Discriminant Analysis

Yasushi Makihara, Yasushi Yagi
The Institute of Scientific and Industrial Research
Osaka University
Ibaraki, Osaka, Japan
Email: {makihara, yagi}@am.sanken.osaka-u.ac.jp

Abstract—Pattern recognition problems often suffer from the larger intra-class variation due to situation variations such as pose, walking speed, and clothing variations in gait recognition. This paper describes a method of discriminant subspace analysis focused on situation cluster pair. In training phase, both a situation cluster discriminant subspace and class discriminant subspaces for the situation cluster pair by using training samples of non recognition-target classes. In testing phase, given a matching pair of patterns of recognition-target classes, posterior of situation cluster pairs is estimated at first, and then the distance is calculated in the corresponding cluster-pairwise class discriminant subspace. The experiments both with simulation data and real data show the effectiveness of the proposed method.

Keywords—linear discriminant analysis; cluster;

I. INTRODUCTION

Discriminant subspace analysis has been one of the most important topics in the pattern recognition research. The most traditional approach is Linear Discriminant Analysis (LDA)[1]. Although LDA have been widely used in many pattern recognition problems such as biometric authentication[2], it is often the case that the difference between the patterns of the same class is larger than that between different classes due to the situation variation. For example, in gait recognition, pose, walking speed, and clothing variations often overwhelm the inter-person variations.

One of solutions to the large intra-class variation is Cluster-based Discriminant Analysis (CDA)[3] or Local Fisher Discriminant Analysis (LFDA)[4]. The method works well if training samples of each recognition-target class for each situation cluster are given, it is, however, difficult to completely collect such samples in practice. For example, it is almost impossible to capture biometric signatures of criminal candidates with situation variations for a surveillance system or an access control system.

On the other hand, it is relatively easy to collect biometric signatures of non recognition-target classes (e.g. members of a laboratory, colleagues of a company, or part time-job or volunteer workers for data acquisition) with situation variations. Actually, several databases with situation variations are open to public in biometrics research areas[5][6].

In addition, we can notice that *the situation cluster recognition problem is easier than the original class recognition problem* in case where the intra-class variation overwhelms

the inter-class variation and where the situations inducing the large intra-class variation are well clustered. This is because small inter-class variation and large intra-class variation in the original problem are converted into small intra-class variation and large inter-class variation in the situation cluster recognition problem.

In such case, once a discriminant subspace for each situation cluster pair, that is, *Cluster-Pairwise Discriminant Analysis (CPDA) subspace*, is constructed with training samples of *non recognition-target classes* in advance, we can choose an appropriate CPDA subspace based on a situation cluster pair recognized for a pair of patterns of *recognition-target classes* in a test phase. Note that the CPDA subspace is more efficient than a generic discriminant subspace for all the situation clusters because intra-class variation only by two situation clusters should be less than that by all the situations. In fact, this kind of pairwise analysis is taken into consideration in several works[7][8][9] and contributes to performance improvement. Note that this paper considers a pair of *situation clusters* while the works[7][8][9] considers a pair of *classes*.

II. TRAINING OF CPDA

A. Dimension reduction

As preprocessing to discriminant analysis, data dimension is reduced in order to avoid curse of dimension and for computational cost reduction. For this purpose, Principal Component Analysis (PCA) is used for vector-form data, and Concurrent Subspaces Analysis (CSA)[10] is used for matrix-form data (2nd order tensor) such as images or even higher order tensors. The reduced dimension is decided so as that the information loss is less than 1% in both cases.

B. Cluster discriminant subspace

Next step is construction of a discriminant subspace of situation clusters. Let a dimension-reduced feature vector of i th training sample be x_i , and a cluster label of i th sample is assumed to be given by a supervised way or an unsupervised way (e.g. by certain clustering methods) as l_i . Next, indice sets of all the training samples, and r th cluster are defined as $I = \{1, \dots, N\}$ and $I^r = \{i | l_i = r, i \in I\}$ respectively. Then, the optimal projection to a discriminant subspace of

the clusters is obtained as

$$\begin{aligned} U^* &= \arg \max_U \frac{\sum_r n^r \|U^T \bar{\mathbf{x}}^r - U^T \bar{\mathbf{x}}\|^2}{\sum_{i \in I} \|U^T \mathbf{x}_i - U^T \bar{\mathbf{x}}^i\|^2} \\ &= \arg \max_U \frac{\text{trace}(U^T S_b U)}{\text{trace}(U^T S_w U)} \end{aligned} \quad (1)$$

where n^r is the number of r th cluster's training samples, and $\bar{\mathbf{x}}^r$ and $\bar{\mathbf{x}}$ are r th cluster average and a total average respectively. Moreover, S_b and S_w are between-cluster and within-cluster covariance matrices respectively. Then, eq. (1) is solved by the following generalized eigen value decomposition.

$$S_b \mathbf{u} = \lambda S_w \mathbf{u} \quad (2)$$

Finally, the optimal projection matrix U^* is obtained as a set of column eigen vectors \mathbf{u} with the largest eigen values λ , and then the dimension reduces feature vector \mathbf{x} is projected to \mathbf{y} in the cluster discriminant subspace as $\mathbf{y} = (U^*)^T \mathbf{x}$.

C. Class discriminant subspaces for cluster pairs

Subsequently, class discriminant subspace for cluster pair is formulated. Let class label of i th training sample and indices sets of c th class be m_i and $I_c = \{i | m_i = c, i \in I\}$ respectively. In addition, let indice set of pair of r th and s th clusters be $I^{(r,s)} = I^r \cup I^s$. Then, the optimal projection to a class discriminant subspace for a cluster pair (r, s) is obtained as

$$\begin{aligned} V^{(r,s)*} &= \arg \max_V \frac{\sum_c n_c^{(r,s)} \|V^T \bar{\mathbf{x}}_c^{(r,s)} - V^T \bar{\mathbf{x}}^{(r,s)}\|^2}{\sum_{i \in I^{(r,s)}} \|V^T \mathbf{x}_i - V^T \bar{\mathbf{x}}^{(r,s)}\|^2} \\ &= \arg \max_V \frac{\text{trace}(V^T S_b^{(r,s)} V)}{\text{trace}(V^T S_w^{(r,s)} V)}, \end{aligned} \quad (3)$$

where $n_c^{(r,s)}$ is the number of training samples of c th class, and $\bar{\mathbf{x}}_c^{(r,s)}$ and $\bar{\mathbf{x}}^{(r,s)}$ are c th class average and a total average respectively. Moreover, $S_b^{(r,s)}$ and $S_w^{(r,s)}$ are between-class and within-class covariance matrices respectively. Note that the above variables are defined for a cluster pair (r, s) . Then, eq. (3) is solved in the same way as the cluster discriminant subspace, and the projected feature vector onto the cluster-pairwise class discriminant subspace is defined as $\mathbf{z}^{(r,s)} = (V^{(r,s)*})^T \mathbf{x}$.

Although LDA is exploited as discriminant analysis in the previous and this sections, DATER[11] can be applied for matrix-form or even higher tensor-form data in the same framework.

III. MATCHING

A. Cluster posterior estimation

Cluster posterior is estimated by non-parametric Probability Density Function (PDF) based on kNN Gaussian kernel. First, let p and \mathbf{y}_p be a probe (input data) and its feature vector in the cluster discriminant space respectively. Then, let $I_p^{k,NN}$ be indices set of k nearest training samples to the

probe feature vector \mathbf{y}_p . Thus, r th cluster posterior $P(r|p)$ is calculated based on the Bayesian rule as

$$P(r|p) = \frac{1}{Z_p} \sum_{i \in I_p^{k,NN} \cap I^r} \exp\left(\frac{\|\mathbf{y}_i - \mathbf{y}_p\|^2}{2\sigma_p^2}\right) \quad (4)$$

$$\sigma_p^2 = \frac{1}{k} \sum_{i \in I_p^{k,NN}} \|\mathbf{y}_i - \mathbf{y}_p\|^2, \quad (5)$$

where Z_p is a distribution function for probability normalization $\sum_r P(r|p) = 1$. As for a gallery (enrolled data) g , the cluster posterior is basically defined in the same way.

In addition, it may be possible in some applications to uniform the cluster in all the galleries. For example, it is relatively feasible to uniform the clothes type in gait-biometrics enrollments. In such case, the common s th cluster posterior of all the galleries is estimated in more robust way as

$$P(s|G) = \frac{1}{Z_G} \sum_{g \in G} \sum_{i \in I_g^{k,NN} \cap I^s} \exp\left(\frac{\|\mathbf{y}_i - \mathbf{y}_g\|^2}{2\sigma(\mathbf{y}_g)^2}\right), \quad (6)$$

where G is set of all the galleries.

Finally, the posterior of cluster pair (r, s) for a probe p and a gallery set G is derived based on observation independence of the probe and the gallery as

$$P((r, s)|p, G) = P(r|p)P(s|G). \quad (7)$$

B. Distance in CPDA subspace

Given feature vectors of a probe p and a gallery g in the class discriminant subspace for cluster pair (r, s) as $\mathbf{z}_p^{(r,s)}$ and $\mathbf{z}_g^{(r,s)}$, the Euclidean distance $d^{(r,s)}(p, g)$ is defined as

$$d^{(r,s)}(p, g) = \|\mathbf{z}_p^{(r,s)} - \mathbf{z}_g^{(r,s)}\|. \quad (8)$$

Then, the distances to all the galleries are z-normalized[12] for better performance in verification scenarios as

$$d^{(r,s)}(p, g; G) = \frac{d^{(r,s)}(p, g) - \mu_d^{(r,s)}(p, G)}{\sigma_d^{(r,s)}(p, G)} \quad (9)$$

$$\mu_d^{(r,s)}(p, G) = \frac{1}{|G|} \sum_{g \in G} d^{(r,s)}(p, g) \quad (10)$$

$$\sigma_d^{(r,s)}(p, G) = \frac{1}{|G|} \sum_{g \in G} (d^{(r,s)}(p, g) - \mu_d^{(r,s)}(p, G))^2. \quad (11)$$

C. Integration

In this section, we describe two methods to integrate a set of distances of all the cluster pairs. The first method is just adopting the distance of the cluster pair with the max posterior

$$D_{max}(p, g; G) = d^{(r^*, s^*)}(p, g; G) \quad (12)$$

$$(r^*, s^*) = \arg \max_{(r, s)} P((r, s)|p, G). \quad (13)$$

The second method is taking expectation of the distances, that is, cluster posterior-expected distance

$$D_{exp}(p, g; G) = \sum_{(r,s)} P((r, s)|p, G) d^{(r,s)}(p, g; G). \quad (14)$$

IV. EXPERIMENTS

A. Simulation data

To confirm the effectiveness of the proposed CPDA, simulation experiments are conducted in this section. First, dimension of original feature vectors, the number of classes and the number of clusters are set as $M = 10$, $N_{class} = 10$, and $N_{cluster} = 5$ respectively. Each cluster center \bar{x}^r is drawn from a uniform distribution in the domain $R_{cluster} = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^M, |x_j| \leq b_{cluster} \forall j\}$. Next, each class center displacement from each cluster center \bar{x}_c is drawn from a uniform distribution in the domain $R_{class} = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^M, |x_j| \leq b_{class} \forall j\}$, and c th class center for r th cluster is defined as $\bar{x}_c^r = \bar{x}^r + \bar{x}_c$. The uniform distribution boundaries are set as $b_{cluster} = 5$ and $b_{class} = 1$ respectively by considering the problem setting where inter-cluster variance is larger than inter-class variance. Then, $N_{sample} = 100$ samples are drawn from Gaussian distribution $\mathcal{N}(\bar{x}_c^r, \sigma^2 I_M)$ with $\sigma = 0.1$ for each class and cluster, and therefore the total number of samples sums up to $N_{cluster} N_{class} N_{sample} = 5000$.

In the problem setting defined in this paper, the training classes and test classes should be separated, the first 5 classes and the last 5 classes are used for training (non-recognition targets) and testing (recognition target) respectively. In addition, gallery samples are drawn only from one cluster in this problem setting, samples in the first cluster and the other clusters of the last 5 classes are assigned to the galleries and the probes respectively. The drawn samples in training and test sets are projected in PCA space as shown in Fig. 1.

Then, the conventional LDA and the proposed CPDA subspace are trained from the training samples respectively, and subsequently the test samples are projected onto the LDA and the CPDA subspace as shown in Fig. 2. As a result, we can see the intra-class variations in the CPDA subspace are well suppressed whereas those in the LDA subspace are diverged.

Finally, performance in a verification scenario is evaluated. The Receiver Operating Characteristics (ROC) curve[12] is the most commonly used in this area, which indicates a tradeoff between false rejection and false acceptance when the receiver changes an acceptance threshold. The ROC curves for PCA, LDA, and CPDA subspace are shown in Fig. 4(a). We can see the CPDA achieves the lowest false rejection rate for all the false acceptance rates.

B. Real data

Another experiment was conducted on real data in a gait verification problem under clothing variation. The dataset

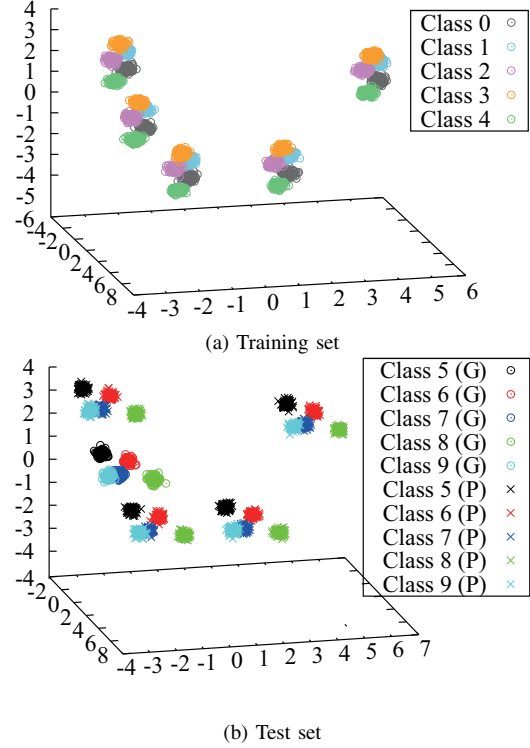


Figure 1. Projection onto PCA space. In legend of (b), (G) and (P) mean gallery and probe respectively. Note that the clusters of gallery and probe are different in (b).

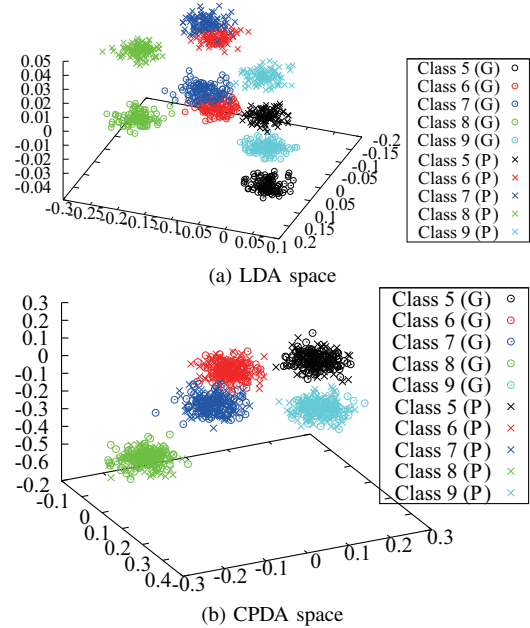


Figure 2. Projection onto discriminant subspaces for a pair of gallery (1st cluster) and a part of probe (2nd cluster) are plotted.

includes 2120 sequences of 68 subjects with at most 32 combinations of clothing types. A part of clothing variation used in this experiment is shown in Fig. 3. From each sequence, a silhouette sequence is extracted by background subtraction-based graph-cut segmentation[13] and it is scaled to 128 by

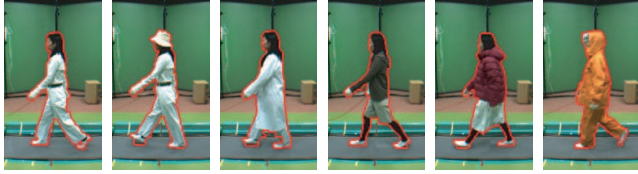


Figure 3. A part of clothing variation

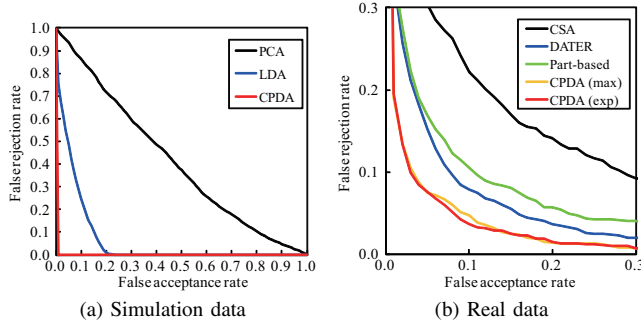


Figure 4. ROC curves.

88 pixel-sized images. Then, the averaged silhouette[14][2] extracted from the silhouette sequence is used as original feature. Because the averaged silhouette is originally matrix-form data, CSA[10] and DATER[11] are exploited instead of PCA and LDA respectively. Finally, the clothing types are clustered into 11 clusters by Ward’s method based on feature distances in the CSA space.

The gait verification performance is evaluated by ROC curve as shown in Fig. 4(b) with several comparison methods, that is, CSA, DATER, and one of the state-of-the-art approaches to the clothing-invariant gait recognition: part-based frequency-domain feature[15]. As a result, the proposed CPDA with two integration methods (distance with max cluster pair posterior and cluster pair posterior-expected distance) outperform the other methods.

V. CONCLUSION

This paper describes a method of discriminant subspace analysis focused on situation cluster pair. In training phase, both a situation cluster discriminant subspace and class discriminant subspaces for the situation cluster pair by using training samples of non recognition-target classes. In testing phase, given a matching pair of patterns of recognition-target classes, posterior of situation cluster pairs is estimated by non-parametric probability density estimation based on k NN Gaussian kernel at first, and then the distance is calculated in the corresponding cluster-pairwise class discriminant subspace. The effectiveness of the proposed method is shown by the experiments both with simulation data and real data in terms of verification performance.

One of future works is further analysis of the effects of the numbers of situation clusters and classes, the ratio of intra-class variation to inter-class variation, on recognition performance.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Research(S) 21220003.

REFERENCES

- [1] N. Otsu, “Optimal linear and nonlinear solutions for least-square discriminant feature extraction,” in *Proc. of the 6th Int. Conf. on Pattern Recognition*, 1982, pp. 557–560.
- [2] J. Han and B. Bhanu, “Individual recognition using gait energy image,” *Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2006.
- [3] T. H. Xue-wen Chen, “Facial expression recognition: A clustering-based approach,” *Pattern Recognition Letters*, vol. 24, pp. 1295–1302, 2003.
- [4] M. Sugiyama, “Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis,” *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, May 2007.
- [5] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer, “The humanid gait challenge problem: Data sets, performance, and analysis,” *Trans. of Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, 2005.
- [6] R. Gross and J. Shi, “The cmu motion of body (mobo) database,” CMT, Tech. Rep., Jun. 2001.
- [7] Y. Li, Y. Gao, and H. Erdogan, “Weighted pairwise scatter to improve linear discriminant analysis,” in *Proc. of the 6th Int. Conf. on Spoken Language Processing*, Oct. 2000, pp. 608–611.
- [8] M. Loog, R. Duin, and R. Haeb-Umbach, “Multiclass linear dimension reduction by weighted pairwise fisher criteria,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 762–766, Jul. 2001.
- [9] M. Robards, J. Gao, and P. Charlton, “A discriminant analysis for undersampled data,” in *Proc. of the 2nd Int. Workshop on Integrating Artificial Intelligence and Data Mining*, vol. 84, 2007, pp. 13–27.
- [10] D. Xu, S. Yan, L. Zhang, H.-J. Z. and Zhengkai Liu, and H.-Y. Shum, “Concurrent subspaces analysis,” in *Proc. of the IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Jun. 2005, pp. 203–208.
- [11] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, “Discriminant analysis with tensor representation,” in *Proc. of the IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Jun. 2005, pp. 526–532.
- [12] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, “The feret evaluation methodology for face-recognition algorithms,” *Trans. of Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [13] Y. Makihara and Y. Yagi, “Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation,” in *Proc. of the 19th Int. Conf. on Pattern Recognition*, Tampa, Florida USA, Dec. 2008.
- [14] Z. Liu and S. Sarkar, “Simplest representation yet for gait recognition: Averaged silhouette,” in *Proc. of the 17th Int. Conf. on Pattern Recognition*, vol. 1, Aug. 2004, pp. 211–214.
- [15] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi, “Clothes-invariant gait identification using part-based adaptive weight control,” in *Proc. of the 19th Int. Conf. on Pattern Recognition*, Tampa, Florida USA, Dec. 2008.