# LEARNING ASSOCIATE APPEARANCE MANIFOLDS FOR CROSS-POSE FACE RECOGNITION

*Xue Chen, Chunheng Wang, Baihua Xiao, Xinyuan Cai*

State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences

## ABSTRACT

Pose variation is a major challenge in face recognition. In this paper, we propose a novel cross-pose face recognition method by learning associate appearance manifolds to model the connection of faces under different poses. The associate manifolds are built on an auxiliary set, in which each identity contains cross-pose face images. The basic assumption is that cross-pose face images from two similar identities can be projected onto similar appearance manifolds by pose-specific transforms. We first associate the input faces with alike identities from the auxiliary set. Then the manifolds of cross-pose faces in the training set are confined close to that of the associate identities in the auxiliary set. Thus, the connection of cross-pose faces is well modeled by the associate appearance manifolds on the auxiliary set. Formally, we formulate the assumption as a manifold-based distance minimization problem, so as to learn the optimal transforms. Experiments on the Multi-PIE dataset demonstrate the effectiveness of the proposed method.

***Index Terms***— cross-pose, face recognition, associate appearance manifolds

## 1. INTRODUCTION

Automatic face recognition systems can achieve high performance under frontal view. However, in real scenarios, face images are generally captured under various poses, which degenerates the performance severely. The difficulty for cross-pose face recognition is that the pose varies in 3D space, while the image captures only 2D appearances. As the pose changes, different visible parts of face appear in the images. It leads to a special phenomenon that faces of different identities with similar poses are more similar than that of the same identity under different poses. The difference brought by variant poses could be larger than that caused by identity changes, making cross-pose face recognition problem very difficult.

To address this problem, many approaches have been proposed [1]. Typically, many researchers use the statistic-based learning methods to seek pose-specific transforms, and then project the images into a common pose-independent subspace. Lin [2] proposed Common Discriminant Feature Extraction (CDEF) to transform samples in different modalities to the common feature space. Sharma [3] introduced Partial Least Squares (PLS) to project faces of different poses to a common linear subspace in which they are highly correlated. Besides, Annan Li [4] applied Canonical Correlation Analysis (CCA) to maximize the intra-individual correlation of samples in the mapping space. Simultaneously, some extensions of these pairwise methods are also developed for multi-view problems, such as Multi-view Discriminant Analysis (MvDA) [5], Multi-view CCA (MCCA) [6].

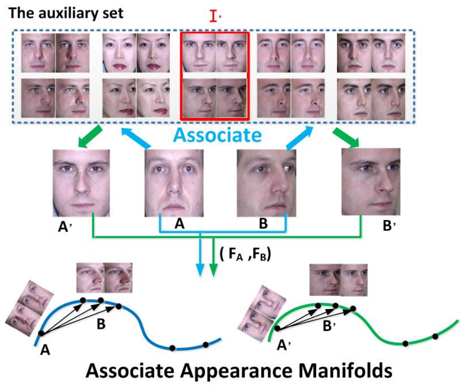

**Fig. 1**. Associate appearance manifolds. $(A, B)$ are faces of identity I with different poses. I' is the alike identity of I, and $(A', B')$ are the associate faces pair of $(A, B)$. Cross-pose faces of the associate identities $(I, I')$ are projected onto similar appearance manifolds by pose-specific transforms $\{F_A, F_B\}$.

Instead of just relying on the sample-to-sample similarity to describe the correlation of intra-class samples, we propose a novel cross-pose face recognition method based on the data distribution structure of cross-pose samples. Concretely, we model the connection of faces from one pose to another by learning associate appearance manifolds (AAMs) on an auxiliary face dataset. As people alike have similar appearance characteristics, it is reasonable to assume that pose-varied face images from two similar identities can be projected onto similar appearance manifolds by pose-specific transforms. First, we associate the input faces with alike identities from the auxiliary set. Then the manifolds of cross-pose faces in the training set are confined close to that of the associate identities in the auxiliary set. Thus, the connection of cross-pose faces is well modeled by the associate appearance manifolds on the auxiliary set. Overview of the proposed method is shown in Fig. 1. When testing, the learned appearance manifold associated with the probe image is used as the reference manifold. Sample, which matches the reference best, is recognized as the probe identity.

## 2. METHODS

In this section, we describe how to recognize cross-pose faces by associate appearance manifolds. The proposed method includes two parts. In the training phase, we first construct identity-coupled face pairs across pose differences, and associate the alike identities from the auxiliary set. Then, the manifolds of face pairs in the training set are projected close to that of the associate identities in the auxiliary set, by pose-specific transforms. In the testing phase, the learned appearance manifold associated with the probe image is used as the reference manifold. Cross-pose face recognition is performed by matching the appearance manifold of the probe face pair with the reference one.

### 2.1. Associate face pairs

In the cross-pose face recognition scenario, images in the gallery and the probe set are from different poses. Formally, assume $T = \{X_i \in X \bigcup Y_i \in Y\}, 1 \leq i \leq C$ be the training set containing C identities. $X = \{X_1, X_2, ..., X_C\}$ is the sample set from pose A, where $X_i = \{x_{i,k} \in \mathbb{R}^{d_A}\}_{k=1}^{N_{X_i}}$ denotes the face images of the $i^{th}$ person, and $N_{X_i}$ is the sample number. $Y = \{Y_1, Y_2, ..., Y_C\}$ holds the corresponding images set $Y_i = \{y_{i,j} \in \mathbb{R}^{d_B}\}_{j=1}^{N_{Y_i}}$ for each person $i$ in $X$, under pose B. $d_A$ and $d_B$ are the sample dimensions. For person $i$ in set $T$, we pair arbitrary image $x_{i,k}$ in $X_i$ with all the images in $Y_i$ to form the identity-coupled face pairs associated with $x_{i,k}$ as $P_{i,k} = \{(x_{i,k}, y_{i,j})\}_{j=1}^{N_{Y_i}}$. In this way, all the identity-coupled face pairs across pose differences in the training set $T$ are constructed as $P = \{P_i\}_{i=1}^{C}$, where $P_i = \{P_{i,k}\}_{k=1}^{N_{X_i}}$ is the face pairs set associated with person $i$.

To construct the associate face pair for arbitrary pair $(x_{i,k}, y_{i,j})$ in set $P$, we first associate one of the input $x_{i,k}$ with alike identity from the auxiliary set. In this paper, rather than constructing an extra auxiliary set, we use the residue set of removing the $i^{th}$ face set $X_i$ from set X, to form the auxiliary set for person $i$ :$\bar{X}_i = \{X_1, ..., X_{i-1}, X_{i+1}, ..., X_C\}$. To associate the most alike identity, we compute the distances between $x_{i,k}$ to all the images of each identity in $\bar{X}_i$, and treat the averaged distance $\bar{d}(x_{i,k}, X_z)$ as the similarity to a specific identity $z$ [7]. Identity holding the minimum distance is identified as the alike identity c. At last, the associate face pair for $(x_{i,k}, y_{i,j})$ are constructed by finding the most similar images for $x_{i,k}$ and $y_{i,j}$ from identity-coupled set $X_c$ and $Y_c$ respectively, which is denoted as $(I(x_{i,k}), I(y_{i,j}))$. As shown in Fig. 1, pair $(A', B')$ from identity $I'$ is the associate face pair of pair $(A, B)$.

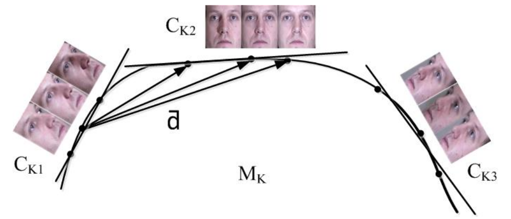

**Fig. 2**. Appearance manifold under varying poses. $M_K$ is the manifold on person K. Each submanifold $C_{K,t}$ on the low-dimensional manifold $M_K$ corresponds to different poses.

### 2.2. Associate appearance manifolds

It is well understood that the set of images of an object under varying viewing conditions can be treated as a low-dimensional manifold in the image space as demonstrated in parametric appearance manifold work [8] or view-based Eigenspace approach [9]. In this paper, we assume that face images of a certain identity under varying poses distribute on a low-dimensional manifold. Specially, images holding similar poses cluster on a local submanifold $C_{K,t}$ of the low-dimensional manifold $M_K = \{C_{K,t}\}_{t=1}^{N_{M_K}}$. An example is shown in Fig. 2. Since images of similar poses have similar appearance, they can be used to reconstruct faces under the same condition linearly. The local submanifolds $C_{K,t}$ consisting of similar faces are approximate linear embeddings.

For cross-pose face recognition scenario, we just consider two local submanifolds $\{C_{K,1}, C_{K,2}\}$ (or $\{C_{K,1}, C_{K,3}\}$) build on cross-pose images set. Concretely, we use the vectors set $\vec{d}$ pointing from samples on one submanifold $C_{K,1}$ to that on another submanifold $C_{K,2}$ to model the distribution characteristic of the joint local submanifolds region $M_{K_{C1,C2}}$, as shown in Fig. 2. Formally, denote the joint submanifolds as $M_{K_{C1,C2}} = \{C_{K,1} \bigcup C_{K,2}\}$, where $C_{K,1} = \{x_{K,i}\}_{i=1}^{N_{X_K}}$ and $C_{K,2} = \{y_{K,j}\}_{j=1}^{N_{Y_K}}$, $x_{K,i}$ and $y_{K,j}$ are images on the submanifolds. Then the distribution characteristic of $M_{K_{C1,C2}}$ is statistically modeled by vectors set $\vec{d^K}$ :

$$\begin{aligned} \vec{d_{i,j}^K} &= x_{K,i} - y_{K,j}, \\ \vec{d^K} &= \{\vec{d_{i,j}^K}\}, \quad i = 1, ..., N_{X_K}; j = 1, ..., N_{Y_K}. \end{aligned} \tag{1}$$

Vectors $\vec{d_{i,j}^K}$ are elements holding size and direction in the feature space. When elements in set $\vec{d^K}$ distribute densely enough in the subpace, the vectors set can rebuild the distribution of the submanifolds $M_{K_{C1,C2}}$ effectively. Therefore, it is rational to use the vectors set for submanifold representation.

As people alike have similar appearance characteristics, it is reasonable to assume that cross-pose faces from two similar identities can be projected onto similar appearance manifolds by pose-specific transforms. Similar kind of approach has been used previously in the literature for other applications like lip-reading [10]. In this paper, we project the manifolds of cross-pose faces in the training set close to that of the associate identities in the auxiliary set, so as to model the connection of cross-pose faces by the learned associate manifolds. Concretely, we confine the vector $\vec{d_{i,j}}$ of each image pair $(x_i, y_j)$ on manifold $M_K$ approximates vector $\vec{d'}_{i,j}$ of the corresponding associate pair $(I(x_i), I(y_j))$ on manifold $M'_K$, so as to make the associate manifolds $\{M_K, M'_K\}$ close with each other.

We denote the transforms for pose A and pose B by $F_a(\theta_A) \in \mathbb{R}^{d' \times d_A}$ and $F_b(\theta_B) \in \mathbb{R}^{d' \times d_B}$, where $d'$ is the mapping dimension of transform matrixes. Given the cross-pose face pairs set $T_{training} = \{(x_{i,k}, y_{i,j}) \in P\}$ and the corresponding associate face pairs set $T_{associate} = \{(I(x_{i,k}), I(y_{i,j})) \in I(P)\}$, we confine the manifolds $\vec{d}$ of cross-pose faces in the training set close to the manifolds $I(\vec{d})$ of the associate identities in the auxiliary set as:

$$J_m(\theta_A, \theta_B) = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N_{X_i}} \sum_{k=1}^{N_{Y_i}} ||\vec{d}^{i}_{j,k} - I(\vec{d}^{i}_{j,k})||^2,$$
$$\vec{d}^{i}_{j,k} = F_a x_{i,j} - F_b y_{i,k}, I(\vec{d}^{i}_{j,k}) = F_a I(x_{i,j}) - F_b I(y_{i,k}), \tag{2}$$

where $N$ is the pair number of associate face pairs. Specially, for multiple-poses recognition tasks, more special constraints should be designed to consider the correlation of multiple-poses local appearance manifolds. We will study this part in our future work.

One problem of the idea above is that it may impose similar associate manifolds on two alike faces holding consistent associate identity. In this case, it is difficult to identify them correctly just via the manifold characteristic. To enhance the discrimination in the mapping space, we add the intra-class compactness regularization into the objective function:

$$J_d(\theta_A, \theta_B) = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N_{X_i}} \sum_{k=1}^{N_{Y_i}} ||F_a x_{i,j} - F_b y_{i,k}||^2. \tag{3}$$

Another important role of the item above is that with compact data distribution, the vectors set $\vec{d}$ could model the distribution characteristic of the manifolds more effectively. To sum up, the proposed model is formulated as follows :

$$\min_{\theta_A, \theta_B} J = J_m(\theta_A, \theta_B) + \alpha * J_d(\theta_A, \theta_B). \tag{4}$$

where $\alpha$ indicates the nonnegative tradeoff parameter.

### 2.3. Solving the optimization model

To solve the problem above with a simply matrix derivation, we reform it in the following way. Let $X = [X_1, ..., X_C]$ collect the images of all the person under pose A, where $X_i = [x_{i,1}, ..., x_{i,N_{X_i}}] \in \mathbb{R}^{d_A \times N_{X_i}}$ is the images of person i. $I(X_i) = [I(x_{i,1}), ..., I(x_{i,N_{X_i}})] \in \mathbb{R}^{d_A \times N_{X_i}}$ collect the corresponding associate images of $X_i$, $1 \leq i \leq C$. Similar denotations are used for image matrix Y under pose B and the corresponding associate matrix I(Y). To construct associate pairs in the form of matrix, we set:

$$\bar{X}_i = [\bar{x}_{i,1}, .., \bar{x}_{i,N_{X_i}}],\ \bar{x}_{i,j} = [x_{i,j}, .., x_{i,j}] \in \mathbb{R}^{d_A \times N_{Y_i}},$$
$$\bar{Y}_i = [Y_i, .., Y_i] \in \mathbb{R}^{d_B \times (N_{Y_i} \times N_{X_i})}. \tag{5}$$

Similar expressions are also used for $I(X_i)$ and $I(Y_i)$. Then, we cast the function in Eq. (4) into a simplified form:

$$\min_{F_A, F_B} J = \frac{1}{N} ||F_a(\bar{X} - I(\bar{X})) - F_b(\bar{Y} - I(\bar{Y}))||_F^2 + \frac{\alpha}{N} ||(F_a \bar{X} - F_b \bar{Y}||_F^2, \tag{6}$$

where $||.||_F^2$ stands for the Frobenius norm of matrix. The gradient descend is used for optimization, in which the gradients $\{\partial J / \partial F_A, \partial J / \partial F_B\}$ are easy to compute.

## 3. RECOGNITION ALGORITHM BY AAMS

For testing, we use the whole training set $T$ as the auxiliary set. To recognize the identity of the probe set $S = \{s_j\}_{j=1}^{N_S}$ under pose A from the gallery $R = \{r_k\}_{k=1}^{N_R}$ under pose B, we first associate the alike face pair $(I(s_j), I(r_k))$ for testing pair $(s_j, r_k)$ from the auxiliary set. Then the manifold on the associate pair $(I(s_j), I(r_k))$ is used as the reference manifold. Cross-pose face recognition is performed by matching the appearance manifolds $\vec{d_{j,k}}$ of the probe face pairs with the reference one $I(\vec{d_{j,k}})$. Concretely, the identity recognition of image $s_j$ is performed as:

$$\mathcal{A} : \hat{k} = arg \min_{k=1,...,N_R} ||\vec{d_{j,k}} - I(\vec{d_{j,k}})||^2 + \alpha * ||\vec{d_{j,k}}||^2, \tag{7}$$

where $\vec{d_{j,k}}$ and $I(\vec{d_{j,k}})$ are defined in the same way as Eq. (2).

## 4. EXPERIMENT

### 4.1. Dataset and Experiment setting

CMU Multi-PIE [11] dataset contains 337 subjects, recorded during four sessions under various poses, illumination and facial expressions. We select 6 images with neutral expression and no flush illuminations of each subject under seven poses $(-45°, -30°, -15°, 0°, 15°, 35°, 45°)$ as the evaluation dataset. The first 231 subjects are used for training, and the next 106 subjects for testing. We used the publicly available labeled locations of facial points [3] to crop the face

| **G\P** | **-45°** | **-30°** | **-15°** | **0°** | **15°** | **30°** | **45°** | Avg |
|---|---|---|---|---|---|---|---|---|
| **-45°** | **100** | **98.4** | **94.2** | 78.5 | 73.4 | 70.5 | 68.0 | 83.3 |
| **-30°** | **98.6** | **100** | **98.9** | **92.4** | 87.7 | 83.4 | 74.9 | 90.8 |
| **-15°** | **96.3** | **98.0** | **100** | **97.4** | **93.9** | 87.1 | 76.7 | 92.8 |
| **0°** | 78.3 | **93.3** | **95.8** | **100** | **99.5** | **93.1** | 86.3 | 92.3 |
| **15°** | 76.3 | 87.5 | **93.7** | **99.8** | **100** | **98.8** | **93.3** | 92.8 |
| **30°** | 74.9 | 79.3 | 83.1 | **93.2** | **99.2** | **100** | **98.5** | 89.7 |
| **45°** | 70.7 | 73.9 | 75.5 | 87.7 | **94.8** | **99.6** | **100** | 86.0 |
| Avg | 85.1 | 90.1 | 91.6 | 92.7 | 92.6 | 90.4 | 85.4 | **89.7** |

**Table 1**. Recognition rate for all pairs of poses (%).

| Methods | FLDA[13] | PLS[3] | CCA[4] | CDEF[2] | AAMs |
|---|---|---|---|---|---|
| Accuracy | 76.7 | 80.1 | 83.2 | 88.8 | **89.7** |

**Table 2**. Mean accuracy of different methods for all possible gallery-probe pairs on Multi-PIE (%).

regions, and then normalize to 32 × 32. Thus, the length of the pixel feature is 1024. Besides, the mapping dimension $d'$ and the tradeoff parameter $\alpha$ are set as 400 and 0.1 respectively. For auxiliary instructions, another common used cross-pose face dataset is the CMU-PIE dataset [12], which just contains 68 subjects. The extra auxiliary set constructed on the training part of the limited CMU-PIE dataset tends to cause serious bias on the associate face pairs and associate appearance manifolds learned. Here, we just experiment on the larger CMU Multi-PIE dataset.

### 4.2. Result and Analysis

Table 1 shows the recognition results for all pairs of poses, and the last row and column give the average accuracy. The bold data in the table show that when the pose difference between the gallery and probe is within ±30°, the proposed method can achieve satisfying accuracy (above 90%). The overall performance of our method is 89.7%, which outperforms the state-of-art methods in Table 2. As training of the proposed method just involves data of two poses, to be fair, we just compare with those pairwise methods. The best reported result is 88.8% from CDEF [2]. This approach just exploits the sample-to-sample distance information to describe the correlation of intra-class samples. In the proposed method, we train the model by incorporating the spatial distribution characteristic of cross-pose samples with the intraclass compactness constraint. The fused space-distance information could model the relation of intra-individuals across pose differences more effectively.

Besides, we also evaluate the effect of the compactness regularization for the proposed method. Fig. 3 shows the results when we set the gallery pose to frontal. Overall, AMMs performs better than that without the regularization (AMMs-without). Particularly, as the probe angle increases, the superiority of AMMs gets bigger comparing with AMMs-without. It indicates that the regularization item promotes the performance of the AMMs significantly, especially when the gap between the gallery and probe pose are large. Actually, when the pose difference of the input pair is large, the data distribution of the manifold on the samples is scattered. The regularization item could enhance the compactedness of the data distribution in the transformed space, which is helpful for effectively modeling the manifold distribution for AMMs.

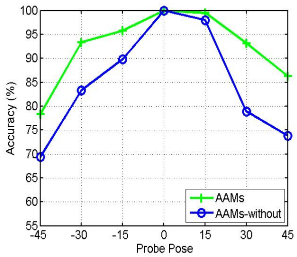

**Fig. 3**. Performance impact of the compactness constraint

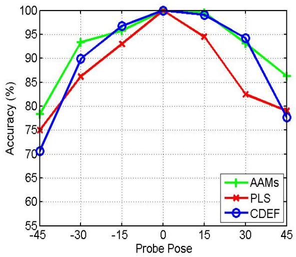

**Fig. 4**. Performance comparison under frontal gallery pose

In addition, Fig. 4 further reports the advantage of our manifold based learning method over the distance based learning methods. We take the classic PLS [3] and CEDF [2] as examples, and perform experiments with the gallery pose as frontal. As seen, the accuracy of PLS is generally worse than the other two. Although CDEF achieves similar results with AMMs when the probe pose is within ±15°, AMMs surpasses it as the probe angle gets large. The result indicates that by exploiting the spatial distribution information, AMMs could better model the connection of cross-pose faces when the different of the input poses is large.

## 5. CONCLUSION

In this paper, we propose a novel approach by learning associate appearance manifolds to deal with the cross-pose face recognition problem. Assuming cross-pose faces from alike identities can be projected onto similar appearance manifolds, the algorithm models the connection of cross-pose faces by the associate appearance manifolds on the auxiliary set. Model learning is performed by confining the manifolds of cross-pose faces in the training set close to that of the associate identities in the auxiliary set. Experiments on the Multi-PIE dataset demonstrate the effectiveness of the proposed method.

## 6. REFERENCES

[1] Xiaozheng Zhang and Yongsheng Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, 2009.

[2] Dahua Lin and Xiaoou Tang, "Inter-modality face recognition," in *ECCV*, pp. 13–26. 2006.

[3] Abhishek Sharma and David W Jacobs, "Bypassing

synthesis: Pls for face recognition with pose, low-resolution and sketch," in *CVPR*, 2011, pp. 593–600.

[4] Annan Li, Shiguang Shan, Xilin Chen, and Wen Gao, "Maximizing intra-individual correlations for face recognition across pose differences," in *CVPR*, 2009, pp. 605–611.

[5] Meina Kan, Shiguang Shan, Haihong Zhang, Shihong Lao, and Xilin Chen, "Multi-view discriminant analysis," in *ECCV*, pp. 808–821. 2012.

[6] Jan Rupnik and John Shawe-Taylor, "Multi-view canonical correlation analysis," in *Conference on Data Mining and Data Warehouses (SiKDD 2010)*, 2010, pp. 1–4.

[7] Qi Yin, Xiaoou Tang, and Jian Sun, "An associate-predict model for face recognition," in *CVPR*, 2011, pp. 497–504.

[8] Hiroshi Murase and Shree K Nayar, "Visual learning and recognition of 3-d objects from appearance," *IJCV*, vol. 14, no. 1, pp. 5–24, 1995.

[9] Alexander Pentland, Baback Moghaddam, and Thad S-tarner, "View-based and modular eigenspaces for face recognition," in *CVPR*, 1994, pp. 84–91.

[10] Ziheng Zhou, Guoying Zhao, and M Pietikainen, "Towards a practical lipreading system," in *CVPR*, 2011, pp. 137–144.

[11] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[12] Terence Sim, Simon Baker, and Maan Bsat, "The cmu pose, illumination, and expression database," *PAMI*, vol. 25, no. 12, pp. 1615–1618, 2003.

[13] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *PAMI*, vol. 19, no. 7, pp. 711–720, 1997.