Smooth Multi-Manifold Embedding for Robust Identity-Independent Head Pose Estimation

Xiangyang Liu[1,2], Hongtao Lu[1], and Heng Luo[1]

[1] MOE-Microsoft Laboratory for Intelligent Computing and Intelligent Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
[2] College of Science, Hohai University, Nanjing, 210098, China
{liuxy,htlu,hengluo}@sjtu.edu.cn

Abstract. In this paper, we propose a supervised Smooth Multi-Manifold Embedding (SMME) method for robust identity-independent head pose estimation. In order to handle the appearance variations caused by identity, we consider the pose data space as multiple manifolds in which each manifold characterizes the underlying subspace of subjects with similar appearance. We then propose a novel embedding criterion to learn each manifold from the exemplar-centered local structure of subjects. The experiment results on the standard databases demonstrates that the SMME is robust to variations of identities and achieves high pose estimation accuracy.

1 Introduction

Head pose estimation from images or videos is a classical problem in computer vision [1]. Robust identity-independent head pose estimation plays a significant role in many human-centered computing applications such as view-independent face detection systems and multi-view face recognition systems.

After neuroscientists emphasized manifold ways of visual perception [2], many researchers indicated that the variations of head pose can be visualized as data points lying on a low-dimensional manifold in the image space of a high dimensionality [3,4]. However, how to extract effective pose features for the low-dimensional manifold, and synchronously ignore appearance variations like changes in identity, scale, illumination, etc [5], remain to be challenging problems due to the nonlinear and high data dimensionality. The focus of this paper is to seek the optimal low-dimensional manifold describing the intrinsical pose variations and to provide a robust identity-independent pose estimator.

The changes of pose images due to identity changes are usually larger than that caused by different poses of same person. Thus, it is difficult to obtain the identity-independent manifold embedding which preserves the pose differences. In this paper, we present a Smooth Multi-Manifold Embedding (SMME) method, which considers the pose data space as multiple manifolds. Each manifold characterizes the underlying subspace of the local structure of subjects with similar appearance. We propose a novel embedding criterion to learn each manifold from

X. Jiang and N. Petkov (Eds.): CAIP 2009, LNCS 5702, pp. 66–73, 2009.

the exemplar-centered local structure of subjects. The embedding method is supervised by both pose and identity information. Each learned manifold with a unique geometric structure is smooth and discriminative. The proposed SMME method aims to provide intra-class compactness and inter-class separability in low-dimensional pose space. For new images of a new subject, we first locate their nearest exemplar, then embed them into the corresponding manifold, and finally decide the pose angle by its k nearest neighbors in the projected subspace.

2 Related Work

The effective manifold learning methods [4,5,6,7,8] for head pose estimation seek a low-dimensional continuous manifold, and new images can then be embedded into these manifolds to estimate the pose. The embedding can be learned by many approaches, such as Locally Embedded Analysis (LEA) [4], and Locality Preserving Projections (LPP) [6]. To incorporate the pose labels that are usually available during training phase, Balasubramanian *et al.* [7] presented a framework based on pose information to compute a biased neighborhood. Yan *et al.* [8] proposed a synchronized manifold embedding method. They all demonstrated their effectiveness for head pose estimation. However, many methods proposed to capture the structure of the pose manifold are local. Thus, they fail to handle new samples without the consistent local information. In addition, they use a single manifold to represent the pose space. In this paper, we use multi-manifold to represent the feature space by a novel embedding method.

Several multi-subspace methods have been proposed in the literature [9,10,11]. Kim *et al.* [9] presented locally linear discriminant analysis for face recognition with a single model image. Vidal *et al.* [10] proposed an algebraic geometric approach to estimate a mixture of subspaces. Tipping *et al.* [11] proposed a mixture model of probabilistic principal component analyzers for face recognition. The parameters of the mixture model are determined using an EM algorithm. They have high computing complexity for the iterative solution methods.

The major contribution of this paper is to introduce the Affinity Propagation (AP) [12] method to obtain local structures of subjects with similar appearance which are used to construct multiple manifolds. Another contribution is the novel formation of the discriminative embedding using the exemplars solved in a closed-form instead of a iterative method.

3 Multi-Manifold Embedding for Head Pose Estimation

Assume that the training data are $X = [x_1^1, x_2^1, \cdots, x_P^1, \cdots, x_1^S, x_2^S, \cdots, x_P^S]_{M \times N}$, where $x_p^s \in R^M$, $s = 1, 2, \cdots, S$, $p = 1, 2, \cdots, P$, S is the number of subjects, and P is the number of poses for a subject α^s, and there are $N = S \times P$ samples in total. The pose angle of the sample x_p^s is denoted as β_p. We aim to seeking a discriminative embedding that mapping the original M dimensional image space into an m dimensional feature space with $m \ll M$.

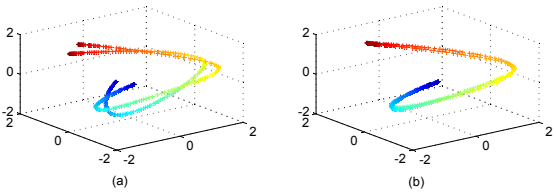


Fig. 1. The 3-dimensional embedding of the pose data by LLE. (a) 2 subjects with dissimilar individual appearance. (b) 2 subjects with similar individual appearance.

3.1 Motivations

The changes of pose images due to identity changes are usually larger than that caused by different poses of same subject. Thus, for head pose estimation, it is crucial to obtain the identity-independent manifold embedding which preserves the pose differences. The SMME method is motivated by two observations: (1) The appearance variations caused by identity lead to translation, rotation and warp changes of the subject's embeddings. Two subjects with similar individual appearance almost lie on a same continuous manifold by Locally Linear Embedding (LLE) [13] shown in Fig. 1-(b). Otherwise, Fig. 1-(a) shows that the embeddings may not be close from two subjects with dissimilar individual appearance. (2) It is difficult to make sure that the pose data lie on a single continuous manifold for the individual variations.

3.2 Smooth Multi-Manifold Embedding

Taking account of the effect caused by the appearance variations from different subjects, we first group subjects in the training data set into clusters (represented by their exemplar), and then seek a discriminative embedding for each cluster supervised by both pose and identity information. Finally, we estimate the pose by the k nearest neighbors in the low-dimensional embedding space.

Clustering Using Affinity Propagation. Frey and Dueck [12] proposed the Affinity Propagation (AP) algorithm which is capable of finding an optimal set of clusters with representative exemplars. Compared with other clustering methods, AP do not preset the number of clusters and has good clustering performance. In our scheme, AP is used to seek the local structures of subjects with similar embeddings in the low-dimensional pose space.

For two head images x_p^s and $x_{p'}^{s'}$, we compute the similarity as follows

$$sim(x_p^s, x_{p'}^{s'}) = -\|x_p^s - x_{p'}^{s'}\|^2. \tag{1}$$

Then, we define the similarity of the two subjects α^i and α^k as

$$s(i,k) = \sum_p sim(x_p^i, x_p^k). \tag{2}$$

The parameter responsibility of AP can be determined experimentally by cross-validation. The output is the clusters $\{X^1, X^2, \cdots, X^K\}$ with the corresponding exemplars $\{x^1, x^2, \cdots, x^K\}$. Later experiments show that each cluster with a exemplar can be used to seek a discriminative embedding.

Embedding Method. For each local structure we seek a low-dimensional embedding to provide intra-class compactness and inter-class separability in the low-dimensional pose subspace. The optimization of the projection is synchronous as follows: (1) Intra-class Compactness: For each pose, the projection minimizes the distances between the embeddings of the exemplar and the other subjects. (2) Inter-class Separability: For each subject, the projection maximizes the distances between the embeddings of the different poses.

To obtain a low-dimensional pose space that is good for pose estimation, it is desirable to minimize the intra-class compactness. We formulate it as the distances between the embeddings of the exemplar and the other subjects for each pose. Namely, we should minimize

$$\sum_{p} \sum_{i \in X^c} \|y_p^i - y_p^c\|^2, \tag{3}$$

where y_p^c is the embedding of the head image x_p^c that is the exemplar of the cluster X^c with the pose angle β_p.

At the same time, we promote the inter-class separability of different poses by maximizing the distances between the embedding of the different poses for each subject. Namely, we maximize

$$\sum_{s} \sum_{i \neq j} \|y_i^s - y_j^s\|^2 T_{ij}, \tag{4}$$

where T_{ij} is a penalty for poses i and j. We introduce a heavy penalty to penalize the poses i and j when they are close to each other, this is given as $T_{ij} = exp(-\|\beta_i - \beta_j\|^2) / \sum_i exp(-\|\beta_i - \beta_j\|^2)$. To combine (3) and (4) simultaneously, we minimize the following objective

$$J = \frac{\sum_p \sum_{i \in X^c} \|y_p^i - y_p^c\|^2}{\sum_s \sum_{i \neq j} \|y_i^s - y_j^s\|^2 T_{ij}}, \tag{5}$$

where J is the objective to seek the embedding y_p^s of the head pose x_p^s.

Fig. 2 (a) shows the intrinsical embeddings from a local structure of four subjects (three subjects denoted by circles and an exemplar denoted by star). The optimization for the projection is to minimize the distances between the exemplar and the other subjects with a same pose and maximize the distances between different poses of a subject. Fig. 2 (b) shows the corresponding embeddings which minimized the distances denoted by the dashed lines and maximized the distances denoted by the solid lines. The objective of the embedding is to generate many pose clusters each corresponding to a specific pose angle.

In this paper, we employ a linear projection approach, namely, the embedding is achieved by seeking a projection matrix $W \in R^{M \times m}$ $(m \ll M)$ such that

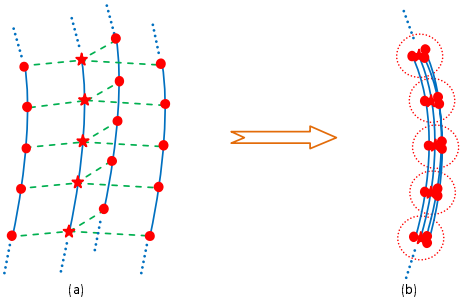


Fig. 2. Illustration of the embedding method. (a) shows the intrinsical embeddings from a local structure of four subjects (three subjects denoted by circles and an exemplar denoted by star). (b) shows the embeddings which minimized the distances denoted by the dashed lines and maximized the distances by the solid lines.

$y_p^s = W^T x_p^s$, where $y_p^s \in R^m$ is the low-dimensional embedding of $x_p^s \in R^M$. Then, W is obtained by the following optimization

$$W^* = \arg\min_W \frac{\sum_p \sum_{i \in X^c} \|W^T x_p^i - W^T x_p^c\|^2}{\sum_s \sum_{i \neq j} \|W^T x_i^s - W^T x_j^s\|^2 T_{ij}}. \tag{6}$$

It is not difficult to see that the objective function can be transformed into

$$W^* = \arg\max_W \frac{Tr(W^T S_2 W)}{Tr(W^T S_1 W)}, \tag{7}$$

where $Tr(\cdot)$ means the trace of a square matrix, and

$$S_1 = \sum_p \sum_{i \in C} (x_p^i - x_p^c)(x_p^i - x_p^c)^T, \quad S_2 = \sum_s \sum_{i \neq j} (x_i^s - x_j^s)(x_i^s - x_j^s)^T T_{ij}. \tag{8}$$

The objective function in (7) can be solved with the generalized eigenvalue decomposition method as $S_2 W_i = \lambda_i S_1 W_i$, where the vector W_i is the eigenvector corresponding to the i-th largest eigenvalue λ_i, and it constitutes the i-th column vector of the projection matrix W.

4 Experiments and Results

The proposed SMME method was validated using the FacePix database [14], which contains 5430 head images spanning $-90°$ to $+90°$ in yaw at $1°$ intervals. We also collected head pose images from the Pointing'04 database [15] for testing. The images were equalized and sub-sampled to 32x32 resolution, and preprocessed by the Laplacian of Gaussian (LoG) filter to capture the edge map that is directly related to pose variations [7].

To evaluate the performance of our system, we use the Mean Absolute Error (MAE) [1] which is computed by averaging the difference between expected pose and estimated pose for all images. To test the generalization ability, we use the leave-one-out strategy [8] (one subject in turn as the testing data and all the remaining subjects for the embedding learning).

4.1 Embedding Space

We use the proposed SMME method on the data sets mentioned above to show the embeddings. Fig. 3-(a) shows two 3-dimensional manifold embeddings from two clusters of 4 subjects with pose variations from $[-75^\circ + 75^\circ]$ at 4° intervals. The result has much better smoothness, intra-class compactness and inter-class separability in the low-dimensional embedding space. And the embedding manifold curves have different geometrical structures and different locations which indicates the multi-manifold representation is benefit for pose estimation.

Fig. 3-(b) shows the distance difference between the image and embedding space for similar poses of the same subject and different subjects with the same pose (We fix the subject 1 with pose 30°, and locate another points by the distance from it). We can see that the distance between images from different subjects with the same pose becomes less than the distance between images from the same subject with similar poses in the low-dimensional embedding space. It indicates that the SMME provides better discriminability for pose estimation.

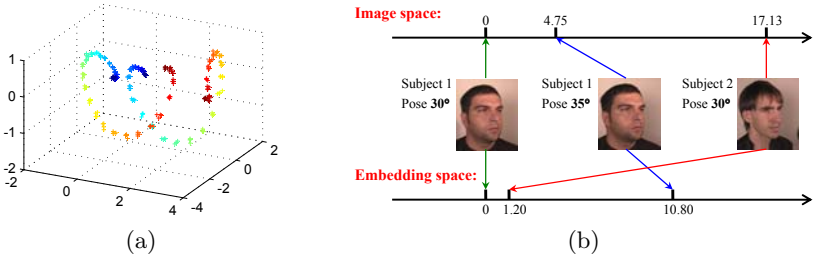
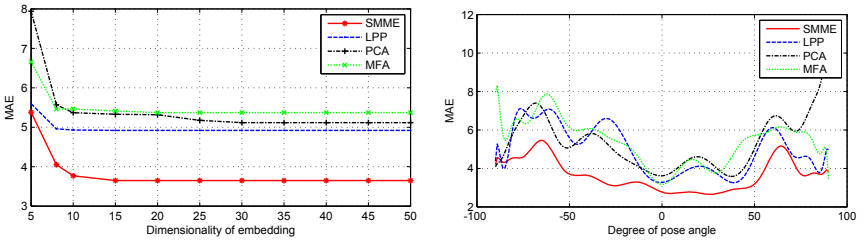


Fig. 3. Illustration of smoothness and discriminability of the embedding space. (a) shows two 3-dimensional manifold embeddings from two cluster. (b) shows the distance difference between the image and embedding space.

4.2 Comparison of SMME with Other Methods

We compare SMME with other pose estimation methods: the global-based PCA method, the local-based manifold learning LPP methods [6] and Marginal Fisher Analysis (MFA) [16] methods. Fig. 4 (a) shows the pose estimation results in different dimensionalities. It shows that the proposed SMME method significantly improves the estimation performance compared to other methods. Fig. 4 (b) shows the MAE with pose variations from $[-90^\circ + 90^\circ]$ at 1° intervals. The result shows that the accuracy of the proposed SMME method is still in general



(a) (b)

Fig. 4. Comparison of our method against other methods. (a) The MAE in different dimensionality. (b)The MAE under different poses.

better than other methods. We notice that the MAE curve of SMME is much more flat than other methods within a relative wide range of the frontal view $[-50° + 50°]$, this implies that SMME is more robust in $[-50° + 50°]$.

4.3 Robustness against Different Identities

In order to test the robustness of SMME against different identities, we use the samples of one subject in turn as the testing data and use all the remaining subjects for embedding learning to compute the MAE of each subjects. The proposed SMME method achieves the average MAE of 3.64° and the variance (for MAE of different subjects) of 1.13 shown in Table 1, which shows that the SMME method can provide more robust and accurate identity-independent head pose estimation than other methods.

Table 1. The MAE of all subjects and the variance of MAE for different subjects

Methods	PCA	MFA	LPP	SMME
MAE	5.32	5.41	4.96	**3.64**
Variance	4.66	4.79	3.21	**1.13**

5 Conclusions

In this paper, we present the SMME method for robust head pose estimation which provides better intra-class compactness and inter-class separability in low-dimensional pose subspace than traditional methods. For identity-independent head pose estimation, we achieved the MAE of about 3° on the standard databases, and even lower MAE can be achieved on larger data sets. In addition, the method has been demonstrated as more robust to individual variations for new identities than the traditional methods. In future, we plan to evaluate the proposed method in terms of feasibility for more complex real world scenarios, and develop auto-adaptive multi-manifold embedding method.

Acknowledgment

This work is supported by the National Laboratory of Pattern Recognition under grant 09-4-1, the National High Technology Research and Development Program of China (No. 2008AA02Z310) and 973 Program 2009CB320900.

References

1. Murphy-Chutorian, E., Trivedi, M.: Head pose estimation in computer vision: a survey. IEEE Transactions on PAMI, 442–449 (2008)
2. Sebastian, H., Lee, D.: The manifold ways of perception. Science 290(12), 2268–2269 (2000)
3. Tenenbaum, J., Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
4. Fu, Y., Huang, T.: Graph embedded analysis for head pose estimation. In: Proc. of International Conference on Automatic Face and Gesture Recognition (2006)
5. Wang, X., Huang, X., Gao, J., Yang, R.: Illumination and person-insensitive head pose estimation using distance metric learning. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 624–637. Springer, Heidelberg (2008)
6. Raytchev, B., Yoda, I., Sakaue, K.: Head pose estimation by nonlinear manifold learning. In: ICPR (2004)
7. Balasubramanian, V., Ye, J., Panchanathan, S.: Biased manifold embedding: a framework for person-independent head pose estimation. In: CVPR (2007)
8. Yan, S., Wang, H., Fu, Y., Yan, J., Tang, X., Huang, T.S.: Synchronized submanifold embedding for person-independent pose estimation and beyond. IEEE Transactions on Image Processing (2008)
9. Kim, T., Kittler, J.: Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. IEEE Transactions on PAMI 27(3), 318–327 (2005)
10. Vidal, R., Ma, Y., Sastry, S.: Generalized principal component analysis (GPCA). IEEE Transactions on PAMI 27(12), 1945–1959 (2005)
11. Tipping, M., Bishop, C.: Mixtures of probabilistic principal component analyzers. Neural computation 11(2), 443–482 (1999)
12. Frey, B., Dueck, D.: Clustering by passing messages between data points. Science 315(514), 972–977 (2007)
13. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
14. Little, D., Krishna, S., Black, J., Panchanathan, S.: A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle. In: ICASSP, vol. 2 (2005)
15. Gourier, N., Hall, D., Crowley, J.: Estimating Face orientation from Robust Detection of Salient Facial Structures. In: VODG, pp. 281–290 (2004)
16. Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. PAMI, 40–51 (2007)