MULTI-MANIFOLD MODELING FOR HEAD POSE ESTIMATION

Xiangyang Liu [1,2] *Hongtao Lu* [1,*] *Wenbin Li* [3]

[1] Department of Computer Science and Engineering, Shanghai Jiao Tong University
Shanghai 200240, China, {liuxy, htlu}@sjtu.edu.cn

[2] College of Science, Hohai University, Nanjing 210098, China
[3] Shanghai Sixth Hospital, Shanghai 200233, China

ABSTRACT

In this paper, we study the identity-independent head pose estimation problem, in order to handle the appearance variations, we consider the pose data lying on multiple manifolds. We present a novel manifold clustering method to construct multiple manifolds each of which characterizes the underlying subspace of some subjects. We first construct a set of n-simplexes of subjects by using the similarity of pose images. Then, we present a supervised method to obtain a low-dimensional manifold embedding for each n-simplex. Finally, we propose the K-manifold clustering method, integrating manifold embedding and clustering, to make each learned manifold with unique geometric structure. The experimental results on a standard database demonstrate that our method is robust to variations of identities and achieves high pose estimation accuracy.

Index Terms— Head pose estimation, Feature extraction, Pattern recognition

1. INTRODUCTION

Head pose estimation plays a significant role in human-machine interaction applications such as view-independent face detection and multi-view face recognition systems [1]. Head pose estimation aims to infer the orientation of a human head from high-dimensional digital imagery or video [2]. It needs to transform a pixel-based representation of a head into a high-level concept of direction [2].

Recently, manifold learning methods have achieved promising results in modeling pose data [3, 4, 5]. The methods seek to define a low-dimensional embedding of the pose data points that preserves some local properties (or geodesic distance) of the high-dimensional pose image set [6, 7]. Due to appearance variations such as changes in identity, scale and illumination, the pose data might lie in multiple different (low-dimensional) manifolds [8]. Thus, many multi-manifold methods have been proposed[9, 10, 11]. Multi-subspace methods are also called Hybrid Linear Modeling (HLM: one linear model for each homogeneous subset of data) [12, 13]. And manifold clustering methods are to classify and parameterize unlabeled data which lie on multiple, intersecting low-dimensional manifolds [6]. Most of them need a mass of data to build the manifold or have high computing complexity for their iterative procedures. Thus, the focus of this paper is to design a new effective multi-manifold method that define multiple low-dimensional manifolds of pose variations and thus to provide a robust identity-independent pose estimator.

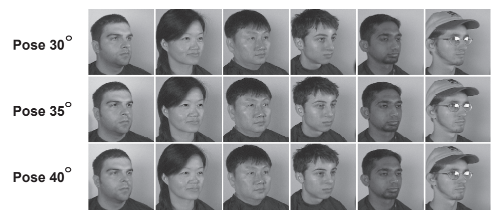


Fig. 1: Head pose images with pose angles $+30°$, $+35°$ and $+40°$ from the FacePix database [14]. (Note that large appearance variations by identity and small variations by pose).

The appearance variations of pose images due to identity changes are usually larger than these caused by different poses as illustrated in Fig. 1. Thus it is difficult to obtain an identity-independent manifold embedding which preserves the pose differences. In this paper, we present a novel manifold clustering method to construct multiple manifolds each of which characterizes the underlying subspace of some subjects. We cover a set of subjects with several clusters each of which contains subjects originating from a separate, single low-dimensional manifold. We use an affinity n-simplex as basic unit to construct the manifold, and then present a supervised method to obtain a low-dimensional manifold embedding for each n-simplex. Finally, we combine the manifold embedding and clustering by the proposed K-manifold clus-

*Correspondent author. This work is supported by the Fundamental Research Funds for the Central Universities(B1020272), Hohai University Funds(2084/409279), NLPR(09-4-1), 863(2008AA02Z310), 973(2009CB320900) and Shanghai Committee of Science and Technology(08411951200,08JG05002).

tering method to make each learned manifold with unique geometric structure. The K-manifold clustering algorithm iteratively computes two nearest n-simplex manifolds, and then constructs a new manifold by the corresponding subjects of the two n-simplexes.

2. MULTI-MANIFOLD FEATURE SPACE

Assume that the training data is $X = \{\alpha^1, \alpha^2, \ldots, \alpha^S\}$, where α^s is the sth subject, and S is the number of subjects. The subject $\alpha^s = [x_1^s, x_2^s, \cdots, x_P^s]$, where P is the number of poses for the subject α^s. x_p^s is an M-dimensional row vector whose elements are taken columnwise from the pth pose image of the sth subject. Thus, there are $N = S \times P$ head pose samples in total. The pose angle of the sample x_p^s is denoted as β_p.

2.1. 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. Our 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 [8]. Two subjects with similar individual appearance almost lie on a same continuous manifold. and 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 [8]. In this paper, we aim to construct a discriminative multi-manifold feature space in which each manifold is an m-dimensional embedding of the original M-dimensional image space.

2.2. Affinity Simplex

We firstly construct a set of n-simplexes of subjects. For two head images x_p^s and $x_{p'}^{s'}$, we first compute the similarity between them as follows

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

Then, we define the similarity of two subjects α^s and $\alpha^{s'}$ by the sum of the similarities as follows

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

Let $k \in N$ and $\alpha \in X$, based on the above definition of the subject similarity, the k-neighborhood of subject α is denoted by $N_k(\alpha)$. Next, we use the affinity-connectivity to define the affinity n-simplex from the simplex [15].

Definition 1 (Affinity-Connectivity) *Let $k \in N$, a subject $\alpha \in X$ is affinity-connected to a subject $\alpha' \in X$ w.r.t k, if α is an element of $N_k(\alpha')$ and α' is an element of $N_k(\alpha)$, formally:*

$$AFFCON(\alpha, \alpha') \Leftrightarrow \alpha \in N_k(\alpha') \cap \alpha' \in N_k(\alpha). \tag{3}$$

Definition 2 (Affinity n-Simplex) *Let $k \in N$ and $n \in N$, a non-empty subset $C \subseteq X$ is called an affinity n-simplex w.r.t k and n, if all subjects in C are affinity-connected and the number of set C is n, formally:*

$$\begin{aligned} &AFFSIM(C) \Leftrightarrow \\ &(1) \quad Connectivity: \forall \alpha, \alpha' \in C, AFFCON(\alpha, \alpha') \\ &(2) \quad n{-}simplex: |C| = n. \end{aligned} \tag{4}$$

2.3. Manifold Embedding

For each affinity n-simplex, we seek a low-dimensional embedding to provide intra-class compactness and inter-class separability in the low-dimensional pose subspace [8].

For the Intra-class compactness, we formulate it as the distances between the embeddings of different subjects for each pose. Namely, we should minimize

$$\sum_p \sum_{i,j} \|y_p^i - y_p^j\|^2, \tag{5}$$

where y_p^i and y_p^j are the embedding of the head image x_p^i and x_p^j with the pose angle β_p. At the same time, we present the inter-class separability of different poses by maximizing the distances between the embeddings of the different poses for each subject. Namely, we maximize

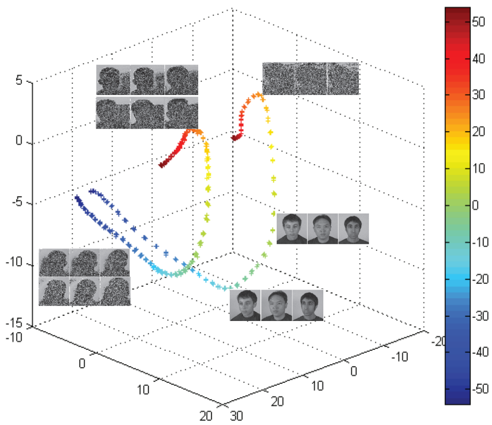
$$\sum_s \sum_{i,j} \|y_i^s - y_j^s\|^2 T_{ij}, \tag{6}$$

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 $T_{ij} = exp(-\|\beta_i - \beta_j\|^2)/\sum_i exp(-\|\beta_i - \beta_j\|^2)$, where β_i is the angle of x_p^i.

2.4. K-Manifold Clustering

In the multi-manifold feature space, each manifold should represent a similar m dimensional embeddings of some subjects. It might be that many manifolds of n-simplexes are similar. This redundancy should be eliminated. To this end, we present a K-manifold clustering algorithm similar to the usual K-means clustering algorithm. While the K-means algorithm basically finds K cluster centers using point to point distance metric, the task here is to find K manifolds using manifold embedding to manifold embedding distance metric in the low-dimensional space. The K-manifold clustering algorithm iteratively computes two nearest n-simplex manifold embeddings, and then construct a new manifold by the corresponding subjects of the two n-simplexes. The new manifold

(a) Two testing images.



(b) 3-D manifold embeddings.

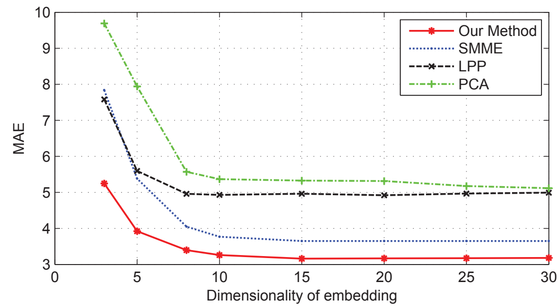
Fig. 2: Illustration of multi-manifold space. (a) Two testing images. (b) 3-D manifold embeddings for the clusters corresponding to the two testing images. (Note that the brightness of the points is corresponding to the pose angle.)

has intra-class compactness and inter-class separability as stated in Section 2.3. The K-manifold clustering method terminates until K clusters are remained $C_t(t = 1, 2, \cdots, K)$, yielding the final results as shown in Fig. 2.

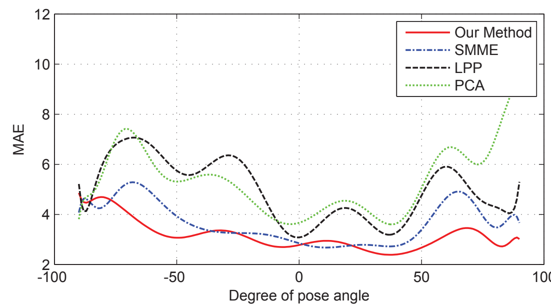
3. EXPERIMENTAL RESULTS AND DISCUSSIONS

Tn this section, the proposed method is validated using the public FacePix database [14], which contains 5,430 head pose images spanning -90° to $+90^\circ$ in yaw at 1° intervals. We use the whole data of FacePix database in our experiments. The images are equalized and sub-sampled to 16x16 resolution, and preprocessed by the Laplacian of Gaussian (LoG) filter to capture the edge map that is directly related to pose variations [3]. We use both the gray and LoG images as the pose features in our experiments.

We evaluate the performance of our system using the Mean Absolute Error (MAE) [2] which is computed by averaging the difference between expected pose and estimated pose for all images. We use the variance (the variance of MAE for different subjects) to evaluate the robustness for identity-independent pose estimation. In order to test the generalization performance, we use the leave-one-out strategy [5] (choose one subject in turn as the testing data and all the remaining subjects for the embedding learning).



(a) The MAE in different dimensionality.



(b) The MAE under different poses.

Fig. 3: Comparison of our method with other methods.

3.1. Multi-Manifold Space

We use the proposed method on the data sets mentioned above to show multi-manifold space. Fig. 2 (a) shows two arbitrary testing images, and Fig. 2 (b) the images in manifold clusters (The left testing image is corresponding to the left manifold embedding of the bottom plot). There are three and six images in the clusters, respectively. The clusters are overlapped and images are weighted in the clusters. The images in the clusters are similar to the testing images in a certain aspect in the whole data set, correspondingly. We construct two 3D manifold embeddings for the clusters corresponding to the two testing images shown in Fig. 2 (b). There are 9 subjects with pose variations from $[-55^\circ + 55^\circ]$ at 3° intervals. It shows that the result has intra-class compactness and inter-class separability in each low-dimensional embedding. And the embedding manifold curves have different geometrical structures and different locations which indicates that the multi-manifold representation is benefit for pose estimation.

3.2. Comparisons with Other Methods

We compare our method with other pose estimation methods: global-based PCA method, local-based manifold learning LPP method [2] and Smooth Multi-Manifold Embedding (SMME) method [8]. The parameters of the methods are set in our experiments as follows: LPP (k=181), SMME (k=6) and our method (n=3 and K=1800). Fig. 3 (a) shows the pose

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

Methods	PCA	LPP	SMME [8]	Our Method
MAE	5.32	4.96	3.64	**3.16**
Variance	4.66	3.21	1.13	**0.98**

estimation results of these algorithms versus different numbers of dimensionality. It shows that the proposed method significantly improves the estimation performance compared to other methods. Fig. 3 (b) shows the MAE with pose variations from $[-90^\circ + 90^\circ]$ at 1° intervals. The result shows that the accuracy of the proposed method is in general better than other methods. We note that the MAE curve of the proposed method is much more flat than other methods within a relative wide range of the frontal view $[-60^\circ + 60^\circ]$, which implies that our method is more robust in $[-60^\circ + 60^\circ]$.

3.3. Robustness against Identities

In order to test the robustness 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 subject. The proposed method achieves the average MAE of 3.16° and the variance of 0.98 as shown in Table 1, which shows that the proposed method provides more robust and accurate identity-independent head pose estimation than other methods. The reason is that the multi-manifold feature space represents the underlying low-dimensional pose space more efficiently and accurately.

4. CONCLUSIONS

In this paper, we proposed a novel multi-manifold algorithm for identity-independent head pose estimation. The proposed multi-manifold method combines the manifold embedding in the K-manifold clustering phase, and has the following characteristics: 1) The embeddings of each manifold are discriminative for different poses; 2) The images in the clusters are weighted. For identity-independent head pose estimation, the proposed method achieved the MAE of 3.16° and the variance of 0.98 on the standard databases. In addition, the proposed method has been demonstrated more robust to individual variations for new identities than the traditional methods. In the future, we plan to theoretically analysis and experimentally evaluate the proposed method in terms of feasibility for more complex real world scenarios.

5. REFERENCES

[1] R. Valenti, Z. Yucel, and T. Gevers, “Robustifying eye center localization by head pose cues,” in *CVPR*, 2009.

[2] Erik Murphy-Chutorian and Mohan Trivedi, “Head pose estimation in computer vision: a survey,” *IEEE Transactions on PAMI*, pp. 442–449, 2008.

[3] V.N. Balasubramanian, J. Ye, and S. Panchanathan, “Biased manifold embedding: a framework for person-independent head pose estimation,” in *CVPR*, 2007.

[4] Y. Fu and T.S. Huang, “Graph embedded analysis for head pose estimation,” in *Proc. of International Conference on Automatic Face and Gesture Recognition*, 2006.

[5] Shuicheng Yan, Huan Wang, Yun Fu, J. Yan, Xiaoou Tang, and Thomas S. Huang, “Synchronized submanifold embedding for person-independent pose estimation and beyond,” *IEEE Transactions on Image Processing*, 2008.

[6] R. Souvenir and R. Pless, “Manifold clustering,” in *ICCV*, 2005, pp. 648–653.

[7] X. Wang, X. Huang, J. Gao, and R. Yang, “Illumination and person-insensitive head pose estimation using distance metric learning,” in *ECCV*, 2008, vol. 2, pp. 624–637.

[8] X. Liu, H Lu, and H Luo, “Smooth Multi-Manifold Embedding for Robust Identity-Independent Head Pose Estimation,” in *Proc. 13th International Conference on CAIP*, 2009.

[9] W. Cao and R. Haralick, “Nonlinear manifold clustering by dimensionality,” in *ICPR*, 2006, vol. 1.

[10] N. Rajpoot, M. Arif, and A. Bhalerao, “Unsupervised learning of shape manifolds,” in *BMVC*, 2007.

[11] A.B. Goldberg, X. Zhu, A. Singh, and Z. Xu, “Multi-Manifold Semi-Supervised Learning,” in *Proc. of International Conference on Artificial Intelligence and Statistics*, 2009.

[12] G. Chen and G. Lerman, “Spectral curvature clustering (SCC),” *International Journal of Computer Vision*, vol. 81, no. 3, pp. 317–330, 2009.

[13] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum, “Estimation of subspace arrangements with applications in modeling and segmenting mixed data,” *SIAM Review*, vol. 50, pp. 413–458, 2008.

[14] D. Little, S. Krishna, J. Black, and S. Panchanathan, “A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle,” in *ICASSP*, 2005, vol. 2.

[15] A. Hatcher, *Algebraic topology*, Cambridge University Press, 2002.