Head Pose Estimation Based on Manifold Embedding and Distance Metric Learning

Xiangyang Liu[1,2], Hongtao Lu[1], and Daqiang Zhang[1]

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

Abstract. In this paper, we propose an embedding method to seek an optimal low-dimensional manifold describing the intrinsical pose variations and to provide an identity-independent head pose estimator. In order to handle the appearance variations caused by identity, we use a learned Mahalanobis distance to seek optimal subjects with similar manifold to construct the embedding. Then, we propose a new smooth and discriminative embedding method supervised by both pose and identity information. To estimate pose of a head new image, we first find its k-nearest neighbors of different subjects, and then embed it into the manifold of the subjects to estimate the pose angle. The empirical study on the standard databases demonstrates that the proposed method achieves high pose estimation accuracy.

1 Introduction

Head Pose Estimation (HPE) from face images or videos is a classical problem in computer vision [1]. It is an integral component of multi-view face recognition systems, human computer interfaces and other human-centered computing applications [2]. Robust and identity-independent head pose estimation remains a challenging computer vision problem.

Face images with varying pose angles were considered to be lying on a smooth low-dimensional manifold in high-dimensional feature space [3,4]. Furthermore, the dimension of this manifold is the degrees of freedom in the variations of head pose [5]. The dimensionality reduction based methods for head pose estimation seek a low-dimensional continuous manifold, and new images can then be embedded into these manifolds to estimate the pose [1]. The fundamental issue is to model the geometry structure of the manifold and produce a faithful embedding for data projection [6]. The classical technique, Principal Component Analysis (PCA), has been used to find the subspace constructed by primary components of training head images for head pose estimation [7,8]. Nevertheless, there is no guarantee that PCA obtains the subspace which is more related to pose variations rather than to appearance variations [1]. The embedding can be directly

H. Zha, R.-i. Taniguchi, and S. Maybank (Eds.): ACCV 2009, Part I, LNCS 5994, pp. 61–70, 2010.

learned by manifold learning approaches, such as Locality Preserving Projections (LPP) [5] and Locally Embedded Analysis (LEA) [4]. To incorporate the pose labels that are usually available during training phase, Balasubramanian *et al.* [9] presented a framework based on pose information to compute a biased neighborhood for each point in the feature space. Based on the framework, Wang *et al.* [10] use Local Fisher Discriminant Analysis (LFDA)[11] to eliminate the variations of illumination in the embedding space. Yan *et al.* [12] proposed a synchronized manifold embedding method by minimizing the distances between each sample and its nearest reconstructed neighbors, and meanwhile maximizing the distances between different sample pairs. They all demonstrated their effectiveness for head pose estimation. However, how to extract effective pose features for the low-dimensional manifold, and synchronously ignore appearance variations like changes in identity, scale, illumination, etc [10], remain to be challenging problems for pose estimation due to the nonlinear and high data dimensionality. The focus of this paper is to seek the optimal low-dimensional manifold embedding describing the intrinsical pose variations and to provide an identity-independent pose estimator.

The changes of pose images due to identity changes are usually larger than that caused by different similar poses shown in Fig. 1. Thus, it is difficult to obtain the identity-independent manifold embedding which preserves the pose differences [13]. In this paper, we propose a manifold embedding method based on Distance Metric Learning (DML) for head pose estimation. We first learn a Mahalanobis distance metric to make the images closer for subjects with similar manifold. Then, we use the learned Mahalanobis distance to seek subjects to construct an optimal embedding. And we propose a new smooth and discriminative embedding method supervised by both pose and identity information. To seek the optimum projection, we minimize the distances between different subjects with the same pose, and maximize the distances between different poses from a subject. The learned manifold with a unique geometric structure is smooth and discriminative. The proposed method aims to provide better intra-class compactness and inter-class separability in low-dimensional pose space than traditional methods. That is, the embeddings in low-dimensional pose space with different poses are kept apart, and the embeddings of different subjects with the same pose are close to each other. For a new image, we first find its k-nearest

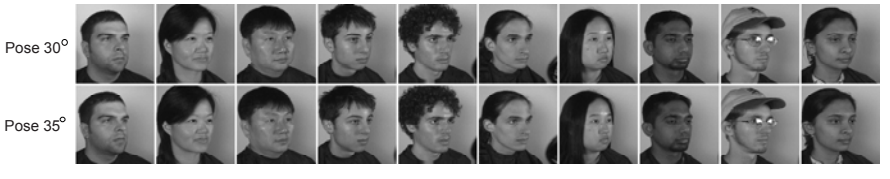

Fig. 1. Head pose images with pose angles $+30^\circ$ and $+35^\circ$ from the FacePix database [14]. (Note that large appearance variations by identity and small variations by pose).