A NEW REPRESENTATION METHOD OF HEAD IMAGES FOR HEAD POSE ESTIMATION

Xiangyang Liu [1,2] *Hongtao Lu* [1 *] *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, {liuxy, htlu, hengluo}@sjtu.edu.cn

[2] College of Science, Hohai University, Nanjing 210098, China

ABSTRACT

In this paper, a discriminative representation method of head images is proposed, which is based on parts and poses for identity-independent head pose estimation. Head images are preprocessed to enhance the facial features and eliminate the identity information by skin color model and Laplacian of Gaussian transform. Then, the preprocessed images are used to construct a eigenpose subspace by a matrix factorization method. The testing head images are represented as the projections of the eigenpose subspace in which we can easily find the decision function for head pose estimation. The proposed representation method evaluated on the standard head pose database and real-time videos achieves higher pose estimation accuracy than other methods.

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

1. INTRODUCTION

Robust and accurate head pose estimation is a classic problem in computer vision because it helps the computer to determine human identity and focus of attention in the scene [1]. It has been widely used in many human-centered computing applications such as view-independent face detection systems [2], multi-view face recognition systems [3].

To satisfy the requirement of the majority of human-machine interaction applications, head pose estimation is supposed to be based on the following design criteria [1]: 1) a desirable representation for head pose estimation should be identity-independent (i.e., the model trained on training data is capable of generalizing on data from unknown identity) [4]; 2) the representation should work well on monocular low-resolution head pose images which is beneficial to increasing the system's efficiency and reduce the cost of the system [5]. Under such conditions, the robustness of the identity-independent head pose estimation should be considered, and the estimation accuracy should be improved because of large-pose variations.

In this paper, we present a head image representation method based on parts and poses for head pose estimation. Parts-based method corresponds better to the intuition of combining parts of head pose in order to create a whole head pose. And poses-based method means projecting a new head pose image to the eigenpose subspace in order to get a maximal response of the corresponding pose. Thus, we propose Poses-based Non-negative Matrix Factorization method (PNMF) to obtain the basic images (eigenposes) which is used to represent the head pose images. Our representation method involves the following three steps. Firstly, head pose images are preprocessed to enhance the facial features and eliminate the identity information by skin color model and Laplacian of Gaussian transform (LoG). Secondly, the preprocessed images are used to construct the eigenpose subspace by the PNMF. Finally, we represent testing head pose images by the projections of the eigenpose subspace in which we can easily estimate the head poses by pattern classification or nonlinear regression methods. The whole representation framework can alleviate the effect of personal information and large pose variations for head pose estimation.

2. RELATED WORK

From literatures [1, 6], we broadly classify the existing methods for head pose estimation into four distinct categories: (1) Template matching methods use the nearest neighbor classification to find the most similar view of a new head pose. (2) Appearance-based methods develop a functional mapping from the image or feature data to a head pose measurement by pattern classification or nonlinear regression tools. (3) Geometric methods [7] use the relative configuration of facial landmarks (such as eyes, mouth, nose tip) to determine poses. (4) Dimensionality reduction methods [4, 6, 8] seek a low-dimensional continuous manifold constrained by the pose variations, and then new images can be embedded into these manifolds and used for template matching or regression.

[*] Correspondent author: Hongtao Lu (htlu@sjtu.edu.cn). This work was supported by the Open Project Program of the National Laboratory of Pattern Recognition (NLPR), 863 Program of China (No. 2008AA02Z310), NSFC (No. 60873133)and 973 Program of China (No. 2009CB320900).

978-1-4244-5654-3/09/$26.00 ©2009 IEEE

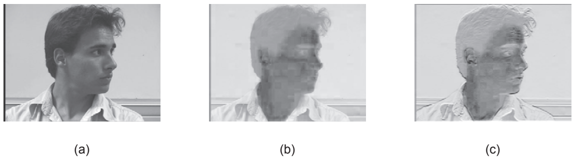

Fig. 1. Illustration of image preprocessing. (a) A gray scale image. (b) Facial features enhanced image by skin color model. (c) Edge enhanced image by the LoG transform.

Geometric methods is considered as local methods, which usually estimate head poses from a set of facial features such as eyes, nose and mouth [7]. However, the features need to be located from high-resolution images, and the difficulties lie in detecting the features with high precision and accuracy [1]. In contrast to local methods, global methods which involve template matching, appearance-based and dimensionality reduction methods use the entire head pose image to estimate head poses. They do not suffer from these problems and have achieved good results for head pose estimation [4, 6].

In this paper, we present a parts-based method for head pose estimation which corresponds to the intuition of combining parts of head pose in order to create a whole one. And we also propose an extension of NMF, called Poses-based Nonnegative Matrix Factorization method (PNMF), which finds the approximate matrix factorization to minimize the KL divergence, and at the same time minimize the distance between the eigenposes and the objective poses.

3. HEAD IMAGE REPRESENTATION

In this section, we will describe our representation method which involves the following steps: The head images are preprocessed to enhance the facial features and eliminate the identity information by skin color model and Laplacian of Gaussian (LoG) transform. Then, the preprocessed images are used to construct the eigenpose subspace. The testing head images are represented as the projections of the eigenpose subspace.

3.1. Head Image Preprocessing

The facial features are much significant to estimate the head pose for us, and the used features should be independent of identity information for identity-independent head pose estimation. Thus, we enhance the facial features by skin color information to stand out more significative facial features than other features (ground features, hair features, etc.). Because pose variation in head pose images is a direct result of geometric transformation and it is irrelevant to texture information, we filter the images with the LoG transform for edge enhancement. This procedure can also eliminate identity information for the identity-independent head pose estimation.

Fig.1 (a) and (b) are respectively an original gray scale image and the enhanced image by the skin color model. After enhanced, the facial features are more remarkable in the head pose image. In the Fig.1, (c) shows the enhanced image by the LoG transform.

3.2. Head Image Representation Methods

Assume that the training images are $\{x_1, x_2, \ldots, x_N\}$, taking values in an M dimensional feature space (i.e. $x_k \in R^{n_1 \times n_2} = R^M$). The representation of the head images aims to find a transformation W mapping the original M dimensional image space into an m dimensional feature space, where m is less than M. The head poses should be easily decided in the low dimensional space by a decision function. Principal Component Analysis (PCA) is an classic global method of dimensionality reduction and used for comparison of our method.

3.2.1. PCA and PCA-eigenposes

PCA uses a linear dimensionality reduction transformation that maximizes the scatter of all projected images and decomposes the images by the basis images (eigenposes). The average head pose is $\overline{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$, and the variation of each head pose from the average is $x'_k = x_k - \overline{x}$. Then, the $W \in R^{M \times m}$ of the total scatter matrix $S_T = \sum_{k=1}^{N} x'_k x'^T_k$ is a set of eigenposes with m is less than M. The original images is reconstructed as linear combinations of the basis images W as $X = WH$. The entries of W and H are of arbitrary signs. In the first rows of Fig. 2, the six PCA-eigenposes explain that the PCA-based head pose representation is a global method from the eigenpose images.

3.2.2. NMF and NMF-eigenposes

Contrary to the PCA, Non-negative Matrix Factorization (NMF) does not allow negative entries in the matrix factors W and H [9]. NMF attempts to find an approximate factorization for $Y = WH \approx X$ that minimizes the divergence D between X and Y subject to $W \geq 0, H \geq 0$ and $W = [w_{ik}]_{M \times m}, \sum_{i=1}^{M} w_{ik} = 1, \forall k$ [9]. The cost function D to be minimized is given explicitly by:

$$D(X, Y) = \sum_{i,j} \left(X_{ij} \log \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij} \right) \quad (1)$$

The cost function is minimized through an iterative process by applying an auxiliary function [9]. When the minimum is found, the basis images (eigenposes) contain parts-based image features in W. The parts-based image features H can just be used to construct the head poses. The second row of Fig. 2 shows six NMF-eigenposes. Each eigenpose is based on parts, and a new head pose image can be represented by the basis images (eigenposes) [9].

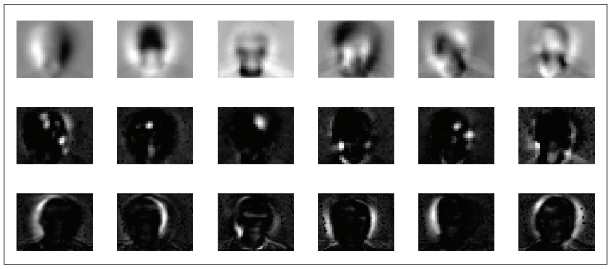

Fig. 2. Illustration of the eigenposes. The PCA-eigenposes are shown in the top row, and the eigenposes obtained by NMF and PNMF are shown in the second and bottom row.

3.2.3. Poses-based NMF (PNMF) and PNMF-eigenposes

The NMF-eigenposes use parts information to represent head pose images. However, the pose information is very important for obtaining the maximum corresponding of the objective poses. Thus, we add poses information to the basic images by imposing additional restrictions to the objective function D in the NMF method. If we have the shape of the objective poses W^{obj}, then they can be used for directing NMF method to get pose-based eigenposes. The cost function D is then:

$$D(X,Y) = \sum_{i,j}\left(X_{ij}\log\frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij}\right) \\ +\lambda\sum_{j,k}\left(W_{jk} - W_{jk}^{obj}\right)^2 \quad , \tag{2}$$

where λ represents a constant for expressing the importance of the additional constraint, and it is set to about 0.5 in the experiments. W^{obj} is the matrix containing the objective poses-based images, and it is approximately taken by the mean of the training head poses in the latter experiment. Taking the derivative with respect to H and W, the gradient algorithm then states:

$$H_{ab} \leftarrow H_{ab}\frac{\sum_{i=1}^{M}(W_{ia}X_{ib})\Big/\sum_{k=1}^{m}(W_{ik}H_{kb})}{\sum_{i=1}^{M}W_{ia}} \tag{3}$$

$$W_{ca} \leftarrow W_{ca}\frac{\sum_{j=1}^{N}(H_{aj}X_{cj})\Big/\sum_{k=1}^{m}(W_{ck}H_{kj})}{\sum_{j=1}^{N}H_{aj} + 2\lambda(W - W^{obj})_{ca}} \tag{4}$$

$$W_{ca} \leftarrow \frac{W_{ca}}{\sum_{j=1}^{M}W_{ja}} \tag{5}$$

From the Eq. (4), we see that the main difference between our new algorithm and the traditional NMF method is the new update rule for W which integrates the poses information.

To summarize, PCA and NMF are respectively a global method and a parts-based method which are reflected in the eigenposes. Fig. 2 shows the eigenposes obtained by the three methods where we can initially see that PNMF provides a new representation method based on different poses and parts. The eigenposes shown in the bottom row are more like the different poses than the others. The first one in the bottom row is right orientation head pose image and the second one is left orientation head pose image.

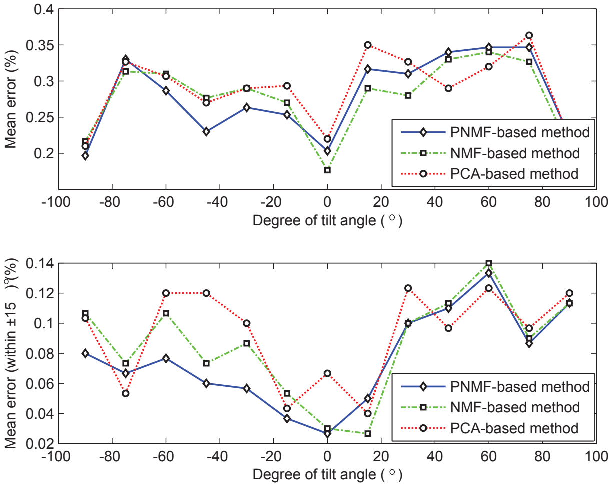

Fig. 3. Mean error with the different degrees of tilt angle.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Experiment with the Standard Databases

We implement the head pose image representation scheme for head pose estimation and present the results on the public head pose database (Pointing'04 database [10]). The Pointing'04 database consists of 15 sets of images. Each set contains 2 series of 93 images of the same person, and the 93 head poses are determined by yaw and tile degrees, which vary from -90° to +90°.

To show the performance of person-independent head pose estimation of low-resolution head pose images, we set the resolution of images to 24×32 pixels. We build each model using the data of 26 subjects, and use the remaining subjects for testing. To evaluate the performance of our system, we use average absolute error which is computed by averaging the difference between expected pose and estimated pose for all images. Precise classification rate and classification within ±15° accuracy [10] are also computed.

The top figure of Fig. 3 shows the mean classification errors of the schemes based on PCA, NMF, HOSVD [11] and PNMF with the different degrees of tilt angle. For large-angle head poses, the error of the PNMF-based scheme is lower than those of the schemes based on PCA, NMF and HOSVD which shows the better performance of PNMF-based representation. In Fig. 3, the bottom figure shows the mean classification errors (within ±15°) of the schemes based on PCA,

Table 1. Results on head pose estimation.

Metric	PNMF	HOSVD[11]	NMF	PCA
Mean yaw err	11.26°	12.90°	12.18°	13.73°
Mean tilt err	12.87°	17.97°	13.51°	14.78°
Classification	50.6%	49.3%	47.8%	45.9%

NMF, HOSVD and PNMF. The error level of the PNMF-based scheme is lower which shows the better stability of PNMF-based representation.

The results are shown as Table 1 based on the same image preprocessing and different metric. The head pose estimators achieve mean yaw error 11.26°, mean tilt error 12.87°, average 50.6% classification accuracy). We can see that the results of the PNMF-based scheme are in all respects better than those of the schemes based on PCA, NMF and HOSVD.

4.2. Experiment with Real-Time Video

We have also performed an experiment to evaluate the performance of our method to estimate the head poses from images in video [12]. The eigenpose subspace is obtained by the two standard data shown in the Section 4.1. The head pose images in video are cropped manually and set to 24×32 pixels, and then used to predict the head poses. Fig. 4 illustrate some images of the example and the corresponding predicted poses which are illustrated by the arrows in the yaw and tilt axes. We can not obtain accurate ground truth to quantitatively evaluate the results, but most of head poses are estimated accurately. The experiment shows that our method performed well and can be used in the real world scenarios.

5. CONCLUSIONS

In this paper, a discriminative representation of head images is proposed for identity-independent head pose estimation. The representation method based on parts and poses has the potential to get better representation for large variational head poses in the scheme. For person-independent head pose estimation, the system achieved average yaw error 11.26° and average tilt error 12.87° on the low-resolution head pose images. The experiment based on real-time videos shows that our method performed well. In future, we plan to evaluate the proposed method in terms of feasibility for more complex real world scenarios.

6. REFERENCES

[1] Erik Murphy-Chutorian and Mohan Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Transactions on PAMI*, pp. 442–449, 2008.

[2] M. Osadchy, Y. Le Cun, and M.L. Miller, "Synergistic face detection and pose estimation with energy model," *Journal of Machine Learning Research*, vol. 8, pp. 1197–1215, 2007.

[3] J. Tu, T. Huang, Y. Xiong, T. Rose, and F. Quek, "Calibrating Head Pose Estimation in Videos for Meeting Room Event Analysis," *ICIP*, pp. 3193–3196, 2006.

[4] S. Yan, H. Wang, Y. Fu, J. Yan, X. Tang, and Thomas S. Huang, "Synchronized Submanifold Embedding for Person-Independent Pose Estimation and Beyond," *IEEE Transactions on Image Processing*, 2008.

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

[6] V.N. Balasubramanian, J. Ye, and S. Panchanathan, "Biased Manifold Embedding: A Framework for Person-Independent Head Pose Estimation," *CVPR*, pp. 1–7, 2007.

[7] J.G. Wang and E. Sung, "EM Enhancement of 3D Head Pose Estimated by Point at Infinity," *Image and Vision Computing*, vol. 25, no. 12, pp. 1864–1874, 2007.

[8] X. Liu, H. Lu, and H. Luo, "Smooth Multi-Manifold Embedding for Robust Identity-Independent Head Pose Estimation," *CAIP*, pp. 66–73, 2009.

[9] D.D. Lee and H.S. Seung, "Learning the Parts of Objects by Non-negative Matrix Factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[10] N. Gourier, D. Hall, and J.L. Crowley, "Estimating Face orientation from Robust Detection of Salient Facial Structures," *Proc. of International Workshop on Visual Observation of Deictic Gestures*, pp. 281–290, 2004.

[11] J. Tu, Y. Fu, Y. Hu, and T. Huang, "Evaluation of Head Pose Estimation for Studio Data," *Lecture Notes in Computer Science*, vol. 4122, pp. 281, 2007.

[12] Michael D. Breitenstein, Daniel Kuettel, Thibaut Weise, and Luc van Gool, "Real-time face pose estimation from single range images," *CVPR*, 2008.

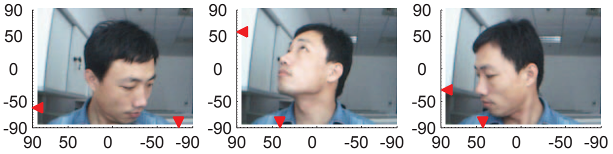

Fig. 4. Pose estimation results based on real-time video. The axes of x and y represent yaw and tilt angle, and the arrows denote the predicted angles by the PNMF method.