

# Three-dimensional face pose estimation based on novel nonlinear discriminant representation

Xinliang Ge, Jie Yang, Tianhao Zhang, Huahua Wang, and Chunhua Du

Jiao Tong University, Institute of Image Processing and Pattern Recognition Shanghai, Number 800 Dongchuan Road, Shanghai 200240, China  
E-mail: xinliang@sjtu.edu.cn

**Abstract.** We investigate the appearance manifold of different face poses using manifold learning. The pose estimation problem is, however, exacerbated by changes in illumination, spatial scale, etc. In addition, manifold learning has some disadvantages. First, the discriminant ability of the low-dimensional subspaces obtained by manifold learning often is lower than traditional dimensionality reduction approaches. Second, manifold learning methods fail to remove the redundancy, such as high-order correlation, among original feature vectors. In this work, we propose a novel approach to address these problems. First, face images are transformed by Gabor filters to obtain a set of overcompleted feature vectors, which can remove intrinsic redundancies within images and provide orientation-selective properties to enhance differences among face poses as well. Second, supervised locality preserving projections (SLPPs) are proposed to reduce dimensionality and obtain the low-dimensional subspace, which has the ability to maximize between-class distance and minimize within-class distance. Finally, the support vector machine (SVM) classifier is applied to estimate face poses. The experimental results show that the proposed approach is effective and efficient. © 2006 Society of Photo-Optical Instrumentation Engineers.  
[DOI: 10.1117/1.2355524]

Subject terms: pose estimation; Gabor filter transformation; locality preserving projections; support vector machine; 3-D face recognition.

Paper 060313LR received Apr. 26, 2006; revised manuscript received Jul. 13, 2006; accepted for publication Jul. 20, 2006; published online Sep. 22, 2006.

## 1 Introduction

Human face pose estimation has a variety of applications, such as face recognition, face tracking, and human-computer interaction (HCI). Due to the inadequacy of the quality of 3-D quantities estimated from 2-D data, it is very complex to estimate face poses from 2-D face images. In addition, many factors exacerbate the problem, for example, illumination conditions, face expressions, spatial scale, etc. More importantly the appearance of the human head can change drastically across different viewing angles, mainly caused by nonlinear deformations during in-depth rotations of the head.<sup>1</sup> Many different approaches have been proposed to solve this problem. Generally, the existing pose estimation methods can be broadly classified into two categories: feature-based<sup>2</sup> and appearance-based methods.<sup>3,4</sup>

There are four major problems to be solved in the existing approaches mentioned before. The first is that the face region must be extracted from a whole face image. It is very difficult to locate the face region from a side or a profile face image. The second is that original face images are normalized manually. However, manual normalization is tedious work and its cost is very high. The third is the difficult problem of extracting the face features accurately. It is especially more difficult to extract features from a side face image than a frontal one. Lastly, face images with varying intrinsic features such as illumination, face pose, and face expression are considered to constitute highly nonlinear manifolds in the high-dimensional observation space. Therefore, some pose estimation systems using linear approaches [for example, principal components analysis (PCA)] will ignore subtleties of manifolds. Manifold learning algorithms are better alternatives. However, the discriminant ability of the low-dimensional subspaces obtained by manifold learning is often lower than those obtained by the traditional dimensionality reduction approaches. Furthermore, the original feature vectors may include high-order correlation, which cannot be removed by manifold learning algorithms. Therefore, a new approach based on manifold learning is proposed to address the four problems mentioned before. In our proposed approach, face images not removed from the background are first transformed by Gabor filters. Then, a novel supervised locality preserving projection (SLPP) is proposed to project Gabor-based data out of the samples into a common low-dimensional subspace. For simplicity, the two combinations of Gabor filters (GF) and SLPP are abbreviated to GF + SLPP. Last, the support vector machine (SVM) classifier is applied to estimate the face pose.

## 2 Proposed Combination Approaches of Gabor Filters and the Supervised Locality Preserving Projection

Gabor filters are particularly appropriate for use in face pose estimation because they incorporate smoothing and can reduce sensitivity to spatial misalignment and illumination change. GWT can also obtain image representations that are locally normalized in intensity and decomposed in spatial frequency and orientation.<sup>5</sup> In addition, Gabor filters can enhance pose-specific face features. Moreover, Gabor filters transform the face images into frequency domain, where unnoticeable information in the spatial domain will become clear. The transformational results of face images do well improving the discriminant ability of SLPP.

In our studies, the system processes face images as follows. A set of Gabor kernels  $h_{mn}(x,y)$  is specified and the original image  $I(x,y)$  is convolved with those kernels at each pixel. The result is a set of 2-D coefficient arrays,

$$W_{mn} = I(x,y) * h_{mn}(x,y), \quad (1)$$

where  $W_{m,n}(x,y)$  is the convolution result corresponding to the Gabor kernel at scale  $m$  and orientation  $n$ .  $*$  denotes the convolution operator.

Since the outputs  $W_{m,n}(x,y)$  consist of different localities, scales, and orientation features, we concatenate all these features into a feature vector  $\mathbf{X}$ . Without loss of generality, assume each output  $W_{m,n}(x,y)$  is a column vector,

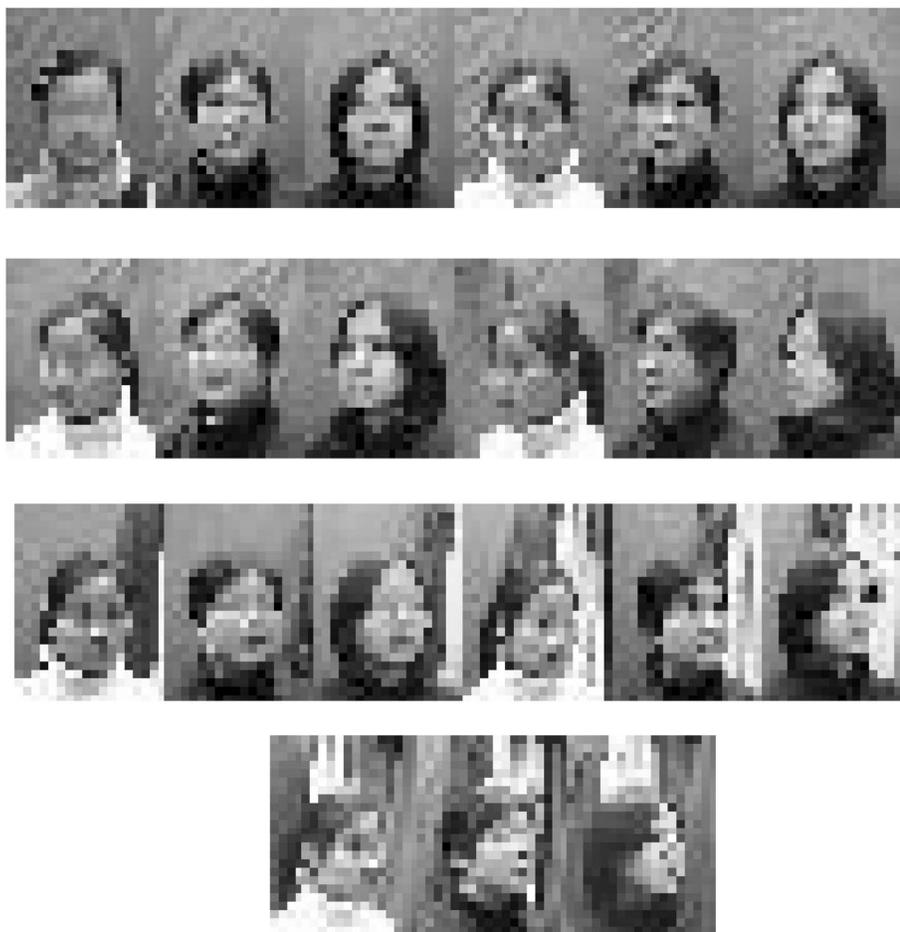


Fig. 1 Some samples of face images in JDL-PEAL face database.

which can be constructed by concatenating the rows (or columns) of the output. Before the concatenation, each output  $W_{m,n}(x,y)$  is down-sampled by a factor  $\rho$  to reduce the dimensionality of the origin vector space. Then, it is nor-

malized to zero mean and unit variance. Let  $W_{m,n}^{\rho}$  denote a normalized output, and then the feature vector  $X^{(\rho)}$  is defined as:

$$X^{(\rho)} = [W_{0,0}^{(\rho)t}, W_{0,1}^{(\rho)t}, \dots, W_{4,7}^{(\rho)t}]^t, \quad (2)$$

where  $t$  is the transpose operator. The feature vector thus encompasses all the output  $W_{m,n}(x,y)$  as important discriminating information.

After high-order information features are extracted by the Gabor filters, an immediate problem is to reduce the dimensionality and uncover the intrinsic low-dimensionality manifold. In this work, we propose a SLPP approach.

LPP seeks a transformation  $W$  to project high-dimensional input data  $X=[x_1, x_2, \dots, x_m]$  into a low-dimensional subspace  $Y=[y_1, y_2, \dots, y_m]$ . The linear transformation  $W$  can be obtained by minimizing an objective function as follows:<sup>6</sup>

$$\min \sum_{i,j} (w^T x_i - w^T x_j)^2 S_{ij}, \quad (3)$$

where  $S_{ij}$  evaluates the local structure of data space. It can be defined as follows:

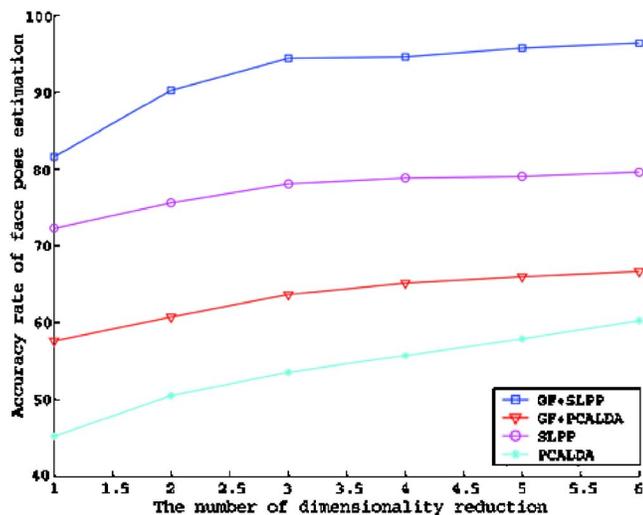


Fig. 2 The influence of the reduced dimension  $d'$  for accuracy rate of face pose estimation.

**Table 1** The accuracy rate (percent) of the combination of dimensionality reduction and SVM classification.  $d=20$  and  $d'=6$ .

Face pose	-45 deg	-30 deg	-15 deg	0 deg	+15 deg	+30 deg	+45 deg
GF+SLPP accuracy rate	96.23	96.53	95.97	97.49	95.93	96.11	96.29
GF+PCALDA accuracy rate	64.14	65.76	69.85	73.52	68.59	66.35	64.62
SLPP accuracy rate	75.23	78.51	78.84	83.85	79.15	77.58	73.39
PCALDA accuracy rate	58.21	59.68	61.18	64.38	61.23	58.92	57.98

$$S_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{t}\right) & \text{if } x_i \text{ and } x_j \text{ are close} \\ 0 & \text{otherwise} \end{cases}, \quad (4)$$

where parameter  $t$  is a suitable constant. The minimization problem can be converted to solving a generalized eigenvalue problem as follows:

$$XLX^T W = \lambda XDX^T W, \quad (5)$$

where  $D_{ii} = \sum_j S_{ji}$  is a diagonal matrix, and  $L = D - S$ . For a more detailed derivation and justifications of LPP, refer to Ref. 6.

The  $d$ -dimensional data from LPP are further mapped into  $d'$ -dimensionality discriminant subspace through the linear discriminant analysis (LDA) algorithm. To minimize the intraclass distances while maximizing the interclass distances of the face manifold, the column vectors of discriminant matrix  $W'$  are calculated by the eigenvectors of  $S_w^{-1}S_b$  associated with the largest eigenvalues,

$$S_w^{-1}S_b W' = \lambda W', \quad (6)$$

where  $S_b$  is the between-class scatter matrix, and  $S_w$  is the within-class scatter matrix. Then the matrix  $W'$  projects vectors in the low-dimensionality face subspace into the common discriminant subspace, which can be formulated as follows:

$$Z = W'Y = W'WXZ \subset R^{d'}, \quad Y \subset R^d, \quad W' \in R^{d' \times d}, \quad (7)$$

where  $Z$  encodes classification information.

### 3 Experimental Results

In this section, we manually selected two collections of face images from the JDL-PEAL face database.<sup>7</sup> They both include 130 subjects, which are selected randomly, each with seven differently posed face images varying intrinsic features such as pose, illumination, and expression. The difference between the two collections is that the first collection is used as a training set while the second one is used as a testing set. In the first collection, all face images were resized to  $24 \times 18$ . Some samples are illustrated in Fig. 1. Before performing the proposed approach, several parameters need to be fixed. First, for the Gabor filters, we chose five scales and eight orientations, and the number of  $\rho$  is 4. Second, the two reduced dimensions  $d$  and  $d'$  of the pro-

posed method are fixed.  $d$  is defined as 20. The reduced discriminant dimension  $d'$  is generally no more than  $L-1$ , where  $L$  denotes the number of face poses.

We compared our proposed GF+SLPP algorithm with PCALDA, GF+PCALDA, and SLPP. For PCALDA, the algorithm is exploited to obtain the subspace in the training set directly. For SLPP, we utilize the SLPP approach without Gabor filters to learn the subspace in the training set. For GF+PCALDA, the approach is similar to the GF+SLPP approach, but the dimensionality reduction approach is replaced by PCALDA.

In the GF+SLPP approach, the reduced discriminant dimension  $d'$  influences the performance of the proposed approach. It can be seen from Fig. 2 that as  $d'$  increases, the GF+SLPP has a higher accuracy rate.

The experimental results with the optimal reduced dimensions are listed in Table 1. It can be seen from Table 1 that the discriminant ability of the SLPP approach is better than the PCA+LDA approach, and the GF+SLPP method achieves the best performance.

### 4 Conclusions

We propose a combination approach of Gabor filters and supervised locality preserving projections. Experimental results show that GF+SLPP has the best performance among all the involved approaches.

#### Acknowledgments

The research is sponsored by the Fundamental Project of the Committee of Science and Technology, Shanghai, under contract 03DZ14015.

#### References

1. B. Gokberk, L. Akarun, and E. Alpaydin, "Feature selection for pose invariant face recognition," *Proc. ICPR*, pp. 306-309 (2002).
2. N. Kruger, M. Potzsch, and C. von der Malsburg, "Determination of face position and pose with a learned representation based on labeled graphs," *Image Vis. Comput.* **15**(10), 741-748 (1997).
3. H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Comput. Vis.* **14**(1), 5-24 (1995).
4. B. Raytchev, I. Yoda, and K. Sakae, "Head pose estimation by non-linear manifold learning," *Proc. ICPR*, pp. 462-466 (2004).
5. C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(5), 572-581 (2004).
6. X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacian faces," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 328-340 (2005).
7. W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large Chinese face database and baseline evaluations," Technical Report on CAS-PEAL.