ROBUST HEAD POSE ESTIMATION USING LGBP

Bingpeng Ma[1,2], Wenchao Zhang[3], Shiguang Shan[1], Xilin Chen[1], Wen Gao[1,2,3]
[1] ICT-ISVISION JDL for Face Recognition, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, 100080, China
[2] Graduate School of the Chinese Academy of Sciences,Beijing,100039, China
[3] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
{bpma, wczhang, sgshan, xlchen, wgao}@jdl.ac.cn

Abstract

In this paper, we introduce a novel discriminative feature which is efficient for pose estimation. The multi-view face representation is based on Local Gabor Binary Patterns(LGBP) and encodes the local facial characteristics in to a compact feature histogram. In LGBP, Gabor filters can extract the feature of the orientation of head and Local Binary Pattern(LBP) can extract the features of facial local orientation. To keep the spatial information of the multi-view face images, LGBP is operated on many subregions of the images. The combination of them can represent well and truly the multi-view face images. Considering the derived feature space, a radial basis function(RBF) kernel SVM classifier is trained to estimate pose. Extensive experiments demonstrate that the facial representation can be effective for pose estimation.

1 Introduction

There has been a significant improvement in face recognition over last ten years. However, the task of robust face recognition is still difficult under pose variation. Pose estimation is a very useful front-end processing for multi-view human face analysis. The methods about pose estimation can be categorized into two main types: model-based approach [1] and appearance-based subspace method [2] [3]. Treating the whole face as a feature vector in some statistic subspaces, appearance-based method can avoid the difficulties of local face feature detection and face modeling in the model-based approach, which has become a popular method recently. But in the subspace method, the distribution of face appearances under variable pose is always a highly non-linear and maybe a twisted manifold, which is hard to be analyzed. Therefore, an exact representation of head poses is very important for the pose estimation.

In this paper, we introduce a novel discriminative feature which is efficient for pose estimation. The multi-view face representation is based on the LGBP [4] operator and consists of encoding the local facial characteristics in to a compact feature histogram. LGBP is actually a representation approach based on multi-resolution spatial histogram combining local intensity distribution with the spatial information. Therefore, it is robust to noises and local image transformations. In addition, LGBP is with much discriminating ability. The idea of using LGBP is motivated by the fact that the shape and orientation of head in the multi-view face images are very important for pose estimation, which means that the good feature for representing the multi-view face images must reflect the orientation of heads. In LGBP, the Gabor filters with 8 different orientations can reflect the orientation of heads in the multi-view face images, and then the LBP [5] operator based on the Gabor features can reflect the local information on the different orientations and different scales. The combination of Gabor and LBP further enhance the representation power of the multi-view face images greatly. And the histogram of the local regions make the discriminative features robust to the misalignment. While regarding classifier design, using these features a RBF kernel SVM classifier is trained to estimate poses. SVM [6] is adopted here since it is well founded in statistical learning theory and has been successfully applied to pose estimation [2]. Extensive experiments on the CAS-PEAL [7] database clearly demonstrate that the LGBP features are effective for pose estimation.

The remaining part of the paper is organized as follows: Section 2 describes the computation of LGBP in multi-view face representation. The pose estimation method based on SVM classifier is presented in Section 3. The experimental results compared with other features and classifiers are shown in Section 4. Some brief conclusions are drawn in Section 5 with some discussions on the future work.

0-7695-2521-0/06/$20.00 (c) 2006 IEEE

2 Feature Description

LGBP is first used for face recognition and attain the impressive result on FERET database [4]. In this approach, a face image is modeled as a "histogram sequence" by concatenating the histograms of the local Gabor magnitude binary pattern maps. And it is impressively insensitive to appearance variations due to lighting, expression, and misalignment. The effectiveness of the LGBP benefits from several aspects including the multi-resolution and multi-orientation Gabor decomposition, the LBP operator, and the local spatial histogram modeling.

The first step of LGBP is to convolute a $64 * 64$ face image $I(x, y)$ with the Gabor filters as follows [8]:

$$G(\mu, \nu) = I(x, y) * \psi_{\mu,\nu}(z) \tag{1}$$

where:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} \mathrm{e}^{(\frac{-\|k_{\mu,\nu}\|^2\|z\|^2}{2\sigma^2})} \left[e^{ik_{\mu,\nu}z} - \mathrm{e}^{\frac{-\sigma^2}{2}}\right] \tag{2}$$

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu}, k_\nu = 2^{-\frac{\nu+2}{2}}\pi, \phi_\mu = \mu\frac{\pi}{8} \tag{3}$$

The processing of facial images by Gabor filters is chosen for its biological relevance and technical properties. We employ a discrete set of 5 different scales, with $\nu = 0, \ldots, 4$, and 8 orientations, with $\mu = 0, \ldots, 7$. And then 40 Gabor Magnitude Pictures(GMPs) can be calculated and the dimension of each GMP is $4096(64 * 64)$.

In the second step, LBP operator operates on each GMP. The original LBP operator labels the pixels of an image by threshold the pixels $f_p(p = 0, \ldots, 7)$ of $3 * 3$ neighborhood with the center value f_c and considering the result as a binary number $S(f_p - f_c)$

$$S(f_p - f_c) = \begin{cases} 1, f_p \geq f_c \\ 0, f_p < f_c \end{cases} \tag{4}$$

Then, by assigning a binomial factor 2^p for each $S(f_p - f_c)$, the LBP pattern at the pixel is achieved as

$$LBP = \sum_{p=0}^{7} S(f_p - f_c)2^p \tag{5}$$

which characterizes the spatial structure of the local image texture. In LGBP, the transform result of $G(\mu, \nu)$ is $\mathbf{LG}(\mu, \nu)$. For the 40 GMPs, there are 40 $\mathbf{LG}$s for each image, and the dimension of each $\mathbf{LG}$ is $4096(64 * 64)$.

To enhance the representation of LGBP, some operations are applied. First, to keep the spatial information of the multi-view face images, LGBP is operated on many subregions of the images. In this case, the new feature description of a full image is written:

$$\mathbf{LGR} = (\mathbf{LG}_{0,0,0}, \cdots, \mathbf{LG}_{\mu,\nu,i}, \cdots, \mathbf{LG}_{4,7,15}); \tag{6}$$

where $\mathbf{LG}_{\mu,\nu,i}$ means the operation of LGBP with orientation μ and scale ν on the i-region of the image. In addition, the histogram information $\mathbf{LGH}_{\mu,\nu,i}$ is extracted from $\mathbf{LG}_{\mu,\nu,i}$ and concatenated into a single histogram sequence $\mathbf{LH}$. We use $\mathbf{LH}$ as our final feature description and we sign the operation of feature description as LGBPH. In our experiment, for a $64 * 64$ image, the region is set as $16 * 16$. And for the number of the bins of the $\mathbf{LGH}_{\mu,\nu,i}$ is 16, the dimension of $\mathbf{LH}$ is $10240(5 * 8 * 16 * 16)$.

3 Application to Pose Estimation

We consider LGBPH features as a facial representation and then build a pose estimation system. A SVM classifier is selected as the classifier in our pose estimation system since it is well founded in statistical learning theory and has been successfully applied to pose estimation.

Given the training samples (multi-view face images) represented by their extracted LGBPH features, a SVM classifier finds the separating hyperplane that has maximum distance to the closest points in the training set. These closest points are called support vectors. To perform a nonlinear separation, the input space is mapped onto a higher dimensional space by function ϕ and the kernel function K is the similarity measurement of the samples in the higher dimensional space. The test data $\mathbf{I}_x$ is classified by the following decision function:

$$F(\mathbf{LH}_x) = Sgn(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{LH}_{I_x}, \mathbf{LH}_{I_i}) + b) \tag{7}$$

Where, α_i is the Lagrange multiplier of the dual optimization problem, y_i is the 1 or -1 depending on whether the training image $\mathbf{I}_i$ is a positive or negative sample, and $K(\cdot, \cdot)$ is a RBF kernel function, $\mathbf{LG}_{\mathbf{I}_i}$ is the LGBP representation of $\mathbf{I}_i$ and b is the parameter of the optimal hyperplane.

Generally, SVM is used for 2-class problems. But, in our approach, the poses of the face images can be viewed as seven classes. So, we use "one against one" approach to solve the k-class problem.

4 Experimental Evaluations

We test our method using the public CAS-PEAL database. The CAS-PEAL database contains seven poses $\{\pm 45°, \pm 30°, \pm 15°, 0°\}$ of 1400 persons. We use a subset including full 1400 images of 200 subjects ranging from 401 to 600. The persons in CAS-PEAL are all Asian.

In experiment, we first label the positions of eyes manually, and then crop the images to $64 * 64$ pixels. Histogram equalization is used for reducing the influence of lighting.

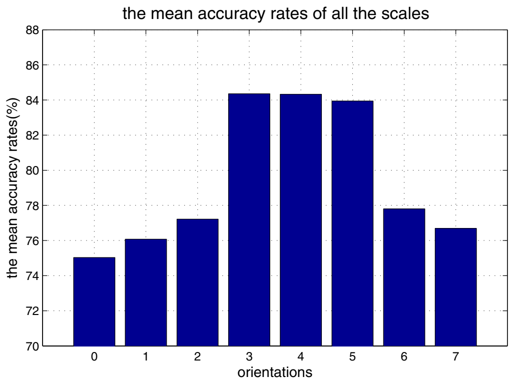

Figure 1. The Mean of Accuracy Rates of All the Scales

Each image is represented by a raster scan vector of the intensity values, and then normalized to be a zero-mean unit-variance vector.

In experiment, we use 3-fold cross-validation in order to avoid over-training. We sort the images by subjects, and divide them into three parts. One part is taken as the testing set and the other two parts are taken as the training set. In this way, the subjects in the training set and the testing set are different. Repeat this until each part has been taken as the testing set. All the results of the experiments are the mean of results of all testing sets.

4.1 The Orientation of LGBP

To show the orientation of the multi-view face images, we use the feature $\mathbf{LG}_{\mu,\nu}$ extracted by each LGBP filter as the input of SVM to estimate poses. As introduced in Section 2, the dimension of $\mathbf{LG}_{\mu,\nu}$ is 4096. The results of pose estimation are shown in Fig. 1. In the figure, the horizontal axis represents the different orientation μ in LGBP, and the vertical axis is the mean accuracy of pose estimation of the orientation μ in all scales. From the figure, the results of the orientation $3, 4, 5(\phi_\mu = \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8})$ are much better than not only the results of other orientations but also the result by using only the LBP operator, which is 83.57%. Therefore, we can get the conclusion that the orientations in LGBP can reflect the shape and orientation of head in the multi-view face images and the LGBP is a better representation for the multi-view face images.

4.2 LGBPH Feature

We use the same SVM classifier to compare the following features: $\mathrm{LBP}_{8,1}^{u_2}$, LBP, wLBP, grey scale feature(G)and LGBPH introduced in Session 2. $\mathrm{LBP}_{8,1}^{u_2}$ is the method using the LBP operator in a $(8, 1)$ neighborhood and using only uniform patterns [5]. And the number of the bins of $\mathrm{LBP}_{8,1}^{u_2}$ is 59. LBP is the standard LBP operator and its bins' number is 256. wLBP is an enhanced histogram of LBP in which an image is divided into many $16 * 16$ regions and then statistic the histogram on each region. For the gray scale feature, PCA is applied for dimension reduction and feature extraction, which can be seen as the baseline of pose estimation. In LGBPH, the grid size of the Gabor operator is set as $32 * 32$. As the dimension of $\mathbf{LH}$ is very high, PCA is used to reduce the dimension to 256, which is the same to the number of bins of LBP and wLBP. The experiment results are shown in Table 1.

Table 1. Pose estimation results of different features

Feature	Accuracy Rate($\%$)
$\mathrm{LBP}_{8,1}^{u_2}$(dim=59)	81.64
LBP(dim=256)	83.57
wLBP(dim=256)	93.14
G (dim=256)	93.64
LGBPH (dim=256)	97.14

From Tab. 1, we can know the result of the $\mathrm{LBP}_{8,1}^{u_2}$ is very low. We attribute this to the unique rotation invariant. In our statistic, the 4.84% histogram information is lost in $\mathrm{LBP}_{8,1}^{u_2}$, which cause the result of $\mathrm{LBP}_{8,1}^{u_2}$ is little descend compared with LBP. The result of LBP is 83.57%, which is lower than the result of gray scale features. As a local feature, the histogram of the LBP feature on the full image lose the spatial information of the pixels. Because both the orientation of head and the spatial information of pixels are very important, we can get the conclusion with briefly the histogram of LBP cannot be used to the pose estimation directly. On the contrary, wLBP can improve results by keeping the spatial information by statistic the histogram on regions.

In all the features, LGBPH get the best accuracy 97.14% of pose estimation in the experiments. The improvements is 13% compared with LBP features and 4% compared with wLBP, which prove that LGBPH can improve the performance of the LBP features greatly. We think the reason of the improvement is that LGBPH is based on the Gabor features, in which the orientations and scales filters can represent the multi-view face image well and truly. And compared with the results of gray feature, which is the baseline in pose estimation, the improvement is 3.5%, which can be attributed the orientations and scales of the Gabor filters in LGBPH. Finally, we think the merits of the LGBPH are: 1) the Gabor filters extract the orientation information, 2) the LBP operator in LGBP extract facial local orientation features, 3) the histogram of LGBP is the statistic of the orien-

Table 2. Pose estimation results of different histogram metrics

Feature	NN			SVM
	D	L	χ^2	
$\mathrm{LBP}_{8,1}^{u_2}$	62.14	24.21	62.36	81.64
LBP	65.29	28.13	64.27	83.57
LGBPH	89.93	82.43	89.64	97.14

tation in the local regions, 4) the regions of LGBPH make LBP can keep the spatial of the multi-view face images.

4.3 SVM Classifier

We use three histogram features: $\mathrm{LBP}_{8,1}^{u_2}$, LBP and LGBPH to compare the performance of SVM classifier with three approaches of histogram matching: histogram intersection(D), log-likelihood statistic(L) and χ^2 statistic. Generally, these three similarity measurement approaches can be taken as the nearest-neighbor(NN) classifiers and they have arrived at the good results in face recognition. The experiment results are shown in Table 2.

From the results, we can know that SVM classifier get the best result in all the features. For the histogram intersection and log-likelihood statistic, they reflect the distribution of the histogram. The results of the D and L denote that the distribution are not exactly to measure the similarity of the multi-view face images. For the multi-view images, there exists two varieties: poses and subjects. And the NN classify can not distinguish exactly that the variety of the multi-view face image is caused by poses or subjects. We attribute the results of SVM to the exact margin. In generally, SVM can find the exact margin when the samples of each class are abundant to represent the true distribute of samples. In our system, there exist about 133 training samples in each class, which satisfy the request of SVM.

5 Conclusion and Future Work

A novel discriminative feature is introduced which is efficient for pose estimation. The representation is based on LGBP and encodes the orientation information of the multi-view face images into an enhanced feature histogram. Considering the proposed representation, we trained a RBF kernel SVM classifier to estimate poses. Extensive experiments clearly show the validity of our approach compared with some other features and classifiers.

The future work should be on the dimensions reduction. The dimension of the LGBPH is the bottleneck in the real-time pose estimation system. And three or less orientation in LGBP may have the same or better results in pose estimation. But, the dimensions are reduced greatly.

6 Acknowledgements

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, “100 Talents Program”of CAS. The authors would also like to thank Fei Yang for revising the draft of the paper. The first author also would like to gratefully acknowledge support from the Natural Science Foundation of China under contract No.60473043.

References

[1] T.F. Cootes, K. Walker and C.J. Taylor, “View-Based Active Appearance Models”. In Proc. 3th Int’l Conf. on Automatic Face and Gesture Recognition, Japan, 1998.

[2] J.N.S. Kwong and S. Gong, “Learning Support Vector Machines for A Multi-View Face Model”. In Proc. of the British Machine Vision Conference 1999, Nottingham, 13-16 September, 1999.

[3] Stan Z. Li, Q. Fu, L. Gu, B. Scholkopf, Y. Cheng and H. Zhang, “Kernel Machine Based Learning For Multi-View Face Detection and Pose Estimation”. In Proc. of 8th IEEE Int’l Conf. on Computer Vision, Vancouver, Canada. July 9-12, 2001.

[4] W. Zhang, S. Shan, W. Gao, X. Chen and H. Zhang, “Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition”. In Proc. of 10th IEEE Int’l Conf. on Computer Vision, Beijing, China, Oct.15-21, 2005.

[5] A. Timo, H. Abdenour and P. Matti, “Face recognition with Local Binary Patterns”. The 8th European Conference on Computer Vision, Prague, May 11-14, 2004.

[6] C. Cortes and V. Vapnik, “Support vector network”, Machine Learning, 20:273:297, 1995.

[7] W. Gao, B. Cao, S. Shan, X. Zhang and D. Zhou, “The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations”. Technical report of JDL, 2004, http://www.jdl.ac.cn/ peal/peal_tr.pdf.

[8] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, R. P. Wurtz, and W. Konen, ”Distortion Invariant Object Recognition in the Dynamic Link Architecture,” IEEE Transactions on Computers, Vol. 42, pp. 300-311, 1993.