LETTER

Fusion of Multiple Facial Features for Age Estimation

Li LU[†a)], ***Student Member and*** **Pengfei SHI**[†], ***Nonmember***

SUMMARY A novel age estimation method is presented which improves performance by fusing complementary information acquired from global and local features of the face. Two-directional two-dimensional principal component analysis ($(2D)^2$PCA) is used for dimensionality reduction and construction of individual feature spaces. Each feature space contributes a confidence value which is calculated by Support vector machines (SVMs). The confidence values of all the facial features are then fused for final age estimation. Experimental results demonstrate that fusing multiple facial features can achieve significant accuracy gains over any single feature. Finally, we propose a fusion method that further improves accuracy.

key words: *age estimation, Gabor wavelets, $(2D)^2$PCA, SVM*

1. Introduction

As an interesting yet challenging problem, age estimation based on face images is a rather new research topic in computer vision. Its applications mainly include commercial data collection (classifying age groups of shopping customers), vending machines refusing to sell alcohol or cigarettes, adult book/video shops and entertainment places (access denied for underage people), etc.

As the earliest attempt to use computer vision techniques for age estimation, the methods proposed by Young et al. [1] was based on cranio-facial changes in feature-position ratios, and on skin wrinkle analysis. Lanitis [2] performed principal component analysis (PCA) on generating facial models which aimed to establish the relationship between the models and the age of the subjects by a quadratic function. Xin Geng et al. [3] modeled the aging pattern, which is defined as the sequence of a particular individual's face images sorted in time order, by constructing a representative subspace. Recently, Guodong Guo [4] learned a low-dimensional embedding of the aging manifold using an appropriate subspace learning method and then designed a locally adjusted robust regressor for learning and prediction of the aging patterns.

However, age estimation is still a challenging problem due to the complex effects caused by living conditions, cosmetics usage, personal specialties, gender differences, and so on. We can see that most computational models of the existing age estimation methods consider only the entire face as a global feature, they do not take into account just what other regions of the face as local features. We propose a novel method for age estimation that combines information from multiple facial features for improving accuracy and robustness. The facial features that we consider are the grayscale image of the face, the Gabor wavelet representation of the face, and the eyes. The Gabor wavelet representations are robust for illumination and expressional variability, so they have been used widely in facial feature modeling in recent years [5]. The eyes are essentially unaffected by beards and mustaches and quite robust to facial expressions and occlusions. Moreover, the area around the eyes was found to be the most significant for the task of age estimation [2]. The idea is to use complementary information for improving overall performance. Finally, we propose a fusion method that further improves accuracy. Two-directional two-dimensional principal component analysis ($(2D)^2$PCA) is used to encode the facial features in a lower dimensional space. Each feature space contributes a confidence value which is calculated by SVMs. The confidence values of all the three features are then fused for final age estimation. The proposed fusion method works quite well and yields a significant improvement in age estimation over that achievable with any single feature.

Manuscript received March 19, 2009.
Manuscript revised May 27, 2008.

[†]The authors are with Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 200240 China.
a) E-mail: lulihappy@sjtu.edu.cn

DOI: 10.1587/transinf.E92.D.1815

2. Global and Local Features

2.1 $(2D)^2$PCA

$(2D)^2$PCA [6] is a transform technique derived from the PCA technique, directly extracts features from image matrices, which is much more efficient than PCA, requiring less memory and having a lower computational cost.

Given a training set $\{X_1, X_2, \cdots, X_N\}$, the ith training image is denoted by an $m \times n$ matrix X_i $(i = 1, 2, \cdots, N)$, and the average image of all training samples is denoted by $\bar{X}$. Construct the image covariance matrix G_1 and G_2:

$$G_1 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^T \tag{1}$$

$$G_2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^T\left(X_i - \bar{X}\right) \tag{2}$$

It has been proven that the optimal value for the projection matrix U is composed by the orthonormal eigenvectors $u_1, \cdots, u_{d_1}$ of G_1 corresponding to the d_1 largest eigenvalues, i.e. $U = [u_1, \cdots, u_{d_1}]$, the optimal value for the pro-

jection matrix V is composed by the orthonormal eigenvectors $v_1, \cdots, v_{d_2}$ of G_2 corresponding to the d_2 largest eigenvalues, i.e. $V = [v_1, \cdots, v_{d_2}]$. For a new image X_{new}, the low-dimensional feature matrix Y can be represented as $Y = UX_{new}V$.

2.2 Gabor Wavelet Representation

The Gabor wavelet representation of a face image is the convolution of the image with a family of Gabor wavelets (filters) [5] which can be defined as follows:

$$\psi_{\mu,v}(z) = \frac{\left\|k_{\mu,v}\right\|^2}{\sigma^2} e^{\left(-\|k_{\mu,v}\|^2 \square \|z\|^2 / 2\sigma^2\right)} \left[e^{ik_{\mu,v}z} - e^{-\sigma^2/2}\right] \quad (3)$$

where μ and v define the orientation and the scale of the Gabor kernels, $z = (x, y)$, $\|\square\|$ denotes the norm operator, and $k_{\mu,v}$ is the wave vector.

For extracting discriminating information of different orientations and scales as much as possible, a bank of Gabor filters with eight orientations, i.e., $\mu \in \{0, \cdots, 7\}$, and five scales, i.e. $v \in \{0, \cdots, 4\}$, is chosen to extract the feature data. Five different scales and eight orientations generate 40 filters. Let $I(x, y)$ be the gray level distribution of a face image, the convolution of image I and a Gabor filter $\psi_{\mu,v}$ is defined as follows:

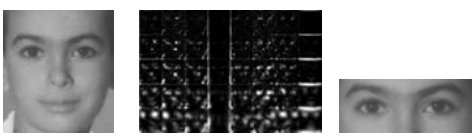
$$O_{\mu,v}(z) = I(z) * \psi_{\mu,v}(z) \quad (4)$$

where $z = (x, y)$, $*$ denotes the convolution operator, and $O_{\mu,v}(z)$ is the convolution result corresponding to the Gabor kernel at orientation μ and scale v. In our work, we use the 2D ensemble Gabor wavelet representation which can be defined as follows:

$$R = \begin{bmatrix} O_{0,0}(z) & O_{0,1}(z) & \cdots & O_{0,4}(z) \\ O_{1,0}(z) & O_{1,1}(z) & \cdots & O_{1,4}(z) \\ \vdots & \vdots & \ddots & \vdots \\ O_{7,0}(z) & O_{7,1}(z) & \cdots & O_{7,4}(z) \end{bmatrix} \quad (5)$$

2.3 Facial Features for Age Estimation

We propose to use global as well as local features. In addition to the entire face image, we consider two other features; namely, the Gabor wavelet representation of the face, and the eyes. Figure 1 shows the extracted facial features for a sample face. The three facial features are high-dimensional and cannot be used directly. We apply $(2D)^2$PCA to each feature for dimensionality reduction.



(a) (b) (c)

Fig. 1 (a) face image, (b) Gabor wavelet representation, and (c) the eyes.

3. The Fusion Approach

3.1 SVMs

SVMs is a supervised learning technique from the field of machine learning applicable to both classification and regression [7]. Consider the training data set $D = \{(x_i, y_i)\}_{i=1}^{l}$ of labeled training samples, $x_i \in R^d$, where d denotes the dimensionality of the training samples, and $y_i \in \{-1, +1\}$ is the associated label. The training vectors x_i are mapped into a high dimensional feature space by the function ϕ. Then SVM finds an optimal linear separating hyperplane that separates all the projected training samples $\phi(x)$ with the maximal margin in this high dimensional feature space. The hyperplane is defined by:

$$f(x) = \sum_{i=1}^{l} \lambda_i y_i k(x, x_i) + b^* \quad (6)$$

where $k(x, x_i) = \phi(x_i)^T \phi\left(x_j\right)$ is called the kernel function, b^* is a bias. To construct an optimal hyperplane is equivalent to finding all the nonzero λ_i and is formulated as a quadratic programming problem with constrains. Based on SVM with RBF kernel $e^{-\gamma\|x_i - x_j\|^2}$ as the binary classifier, one can extend SVM for probability estimates.

3.2 Weighted Summation Fusion

It is often difficult to precisely make out a stranger's age. More applicably, we divide the age samples into 5 categories which range from 1~9, 10~19, 20~39, 40~59, and over 60 years old, implying that our estimation problem is translated into five-class pattern recognition problem. We apply SVMs on facial feature m ($m = 1, 2, \cdots, M$, where M is the number of facial features) to obtain the probability, n_i^m, that the testing observation is in class $i (i = 1, 2, \cdots, I$, where I is the number of age ranges). n_i^m is termed as the confidence value.

We experimented with four different fusion approaches, namely simply-summation, product, maximum, weighted summation. The first three are well-know fusion approaches; the last one is proposed in this paper and it takes into account the performance of individual facial features in weighting their contributions. Given confidence value n_i^m, the fused score for class i is denoted as f_i.

Simply-summation (SS): $f_i = \sum_{m=1}^{M} n_i^m, \ \forall i$

Product (PRO): $f_i = \prod_{m=1}^{M} n_i^m, \ \forall i$

Maximum (MAX): $f_i = \max\left(n_i^1, n_i^2, \cdots, n_i^M\right), \ \forall i$

Weighted summation (WS): Weights are assigned to the individual features based on the reliability of facial features

on age estimation. The distribution of the confidence values of a facial feature contains information about the reliability of that facial feature's decision. For the facial feature m, if the highest ranked class receives a high confidence value and all of the other classes receive relatively low values, or if two classes receives relatively high confidence values and all the other classes receive relatively low values, then the confidence level is high. Conversely, if all the classes receive similar scores, the confidence is low. For a facial feature m, we have the set of ranked confidence values $\left\{n^m_{r_i}\right\}_{i=1\cdots I}$, where I is the number of age ranges. We define γ as the reliability measure which is the difference between the two highest ranked confidence values and the doubled third ranked confidence value, and normalized by the mean value.

$$\gamma^m = \frac{n^m_{r_1} + n^m_{r_2} - 2n^m_{r_3}}{n^m_{mean}} \tag{7}$$

where r_1, r_2 and r_3 are the subject classes achieving the highest, second and third ranks, respectively, m denotes the number of facial features, and the mean is calculated over all the values of $n^m_{r_i}$. Weights are proportional to γ as follows

$$w^m = \frac{1}{\sum_{m=1}^{M} \gamma^m} \cdot \gamma^m \tag{8}$$

Note that $0 \le w^m \le 1,\ \forall m$, and $\sum_{m=1}^{M} w^m = 1$. The WS fused score for class $i$is calculated as

$$f_i = \sum_{m=1}^{M} w^m n^m_i,\ \forall i \tag{9}$$

4. Experimental Results

To verify the proposed age estimation method, we constructed two galleries using face images from the FG-NET Aging Database [8] and the CAS-PEAL database [9], and pictures collected from World Wide Web. In FG-NET, the ages are distributed highly unevenly in the ranges: 1-69, face images in the ranges of 20-39, 40-59, and above 60 years old is not enough for experiments. We therefore used images from the FG-NET as well as the CAS-PEAL and World Wide Web to construct Gallery I as shown in Table 1. Gallery I contains face images of Caucasian and Asian descent. For comparison, images from the CAS-PEAL and World Wide Web are used to construct Gallery II which contains only Asian descent. The face images used in our experiments are mostly frontal faces with some variations in pose, illumination and facial expressions.

All the face images were warped to the same scale, orientation and position. Histogram equalization was applied to the extracted face images to normalize for different lighting conditions. The required facial features were cropped with reference to the eye locations.

The average classification accuracy was estimated with five-fold cross validation (CV) — i.e., a five-way data set split, with 4/5th used for training and 1/5th used for testing, with four subsequent nonoverlapping rotations. For each facial feature, the average size of the training set is 240 and the average size of the test set is 60. We used a SVM classifier with RBF kernel. The SVMs come from LIBSVM [10] and all parameters were selected using cross validation via parallel grid research. Figure 2 (a) illustrates the classification accuracy of different features viz., face, the Gabor wavelet representation of face image (GW), eyes, and the WS fusion method case under different dimensions. Figure 2 (b) shows the classification accuracy of different fusion methods. We verified the proposed age estimation method on Gallery II with the same experimental methodology as we used on Gallery I. The experimental results are shown in Fig. 3.

As shown in Fig. 2 (a) and Fig. 3 (a), all three facial features produced high classification rates, indicating each of them contains a high amount of age information. The Gabor wavelet representation of face image yields the highest classification rate of 85.4% on Gallery I and 86.3% on Gallery II. The reasons may be that the Gabor wavelet representations of face image can extract the local facial fea-

Table 1 Description of two galleries: (a) Gallery I, and (b) Gallery II.

Age ranges	FG-NET (No. data)	CAS-PEAL (No. data)	World Wide Web (No. data)	Total (No. data)
1-9	300	0	0	300
10-19	300	0	0	300
20-39	200	100	0	300
40-59	50	10	240	300
60-	5	4	291	300

(a)

Age ranges	CAS-PEAL (No. data)	World Wide Web (No. data)	Total (No. data)
1-9	0	300	300
10-19	0	300	300
20-39	300	0	300
40-59	10	290	300
60-	4	296	300

(b)

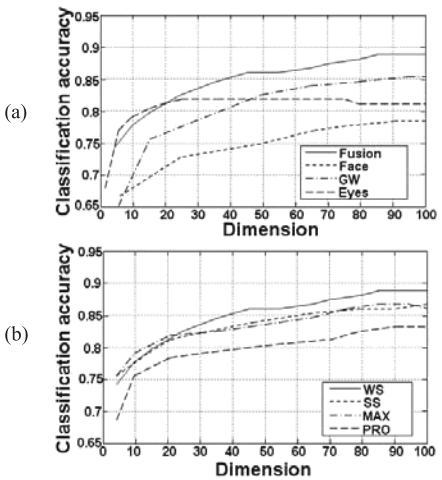


Fig. 2 Experiments on Gallery I.

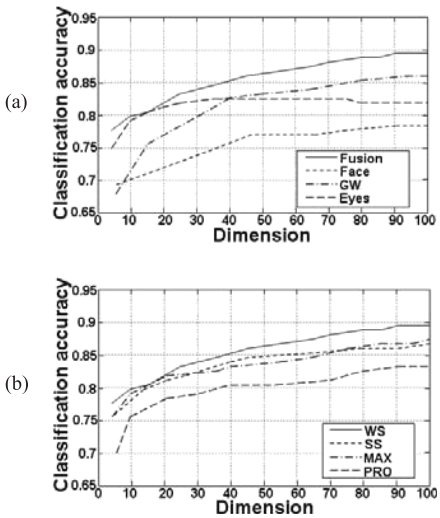


Fig. 3 Experiments on Gallery II.

ture like the analysis of face's wrinkles and shapes for classification, and are robust for illumination and expressional variability. Under the same dimensions of feature vectors, the eyes obtain better accuracy than the entire face image, one reason may be that the eyes essentially unaffected by beards and mustaches, and quite robust to facial expressions. The proposed WS fusion approach, i.e. weighted summation of the classification results of three facial features, obtains the highest correct classification rate of 88.5% on Gallery I and 89.6% on Gallery II. The results demonstrate that fusing multiple facial features can achieve significant accuracy gains over any single feature.

We have experimented with several fusion methods for three facial features, including simply summation, maximum, production rules, and the proposed weight summation method. Figure 2 (b) and Fig. 3 (b) shows the effect of each fusion method for fusing facial features. We see that the weight summation fusion method generally performed better than the other three (SS, PRO and MAX).

Although Gallery I contains face images of different descents, we found the classification accuracy results of Gallery I are close to that of Gallery II. The general accuracy on Gallery II is slightly better than that on Gallery I. This may show that the aging process is common for human beings.

5. Conclusions

We have proposed a novel methodology that combines information gathered from multiple facial features, namely, the face, the Gabor wavelet representation of face, and the eyes, for robust and accurate age estimation. Each feature space contributes a confidence value which is calculated by SVMs. The confidence values of all the facial features are then fused for final age estimation. The experimental results demonstrate that fusing multiple facial features can achieve higher accuracy over any single feature, the proposed weighted summation fusion method outperforms the traditional fusion methods and can further improve accuracy.

Acknowledgements

This work is supported by National Natural Science Foundation of China under grant No.60775009.

References

[1] Y.H. Kwon and N. da V. Lobo, "Age classification from facial images," Computer Vision and Image Understanding, vol.74, no.1, pp.1–21, 1999.
[2] A. Lanitis, "On the significance of different facial parts for automatic age estimation," 14th International Conference on Digital Signal Processing, 2002.
[3] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.12, Dec. 2007.
[4] G. Guo, Y. Fu, C.R. Dyer, and T.S. Huang, "Image-based human age estimation by manifold learning locally adjusted robust regression," IEEE Trans. Image Process., vol.12, no.7, pp.1178–1188, 2008.
[5] L. Wang, Y. Li, C. Wang, and H. Zhang, "2D gaborface representation method for face recognition with ensemble and multichannel model," Image Vis. Comput., vol.26, no.6, pp.820–828, 2008.
[6] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "An approach for directly extracting features from matrix data and its application in face recognition," Neurocomputing, vol.71, no.10-12, pp.1857–1865, 2008.
[7] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer Verlag, New York, 1995.
[8] The FG-NET Aging Database [Online], Available: http://www.fgnet.rsunit.com/, 2002.
[9] W. Gao, B. Cao, S. Shan, D. Zhou, et al., "The CAS-PEAL large-scale Chinese face database and Baseline evaluations," technical report of JDL, 2004.
[10] C.C. Chang and C.J. Lin, "LIBSVM: A library for support vector machines," http://www.csie.ntu.edu.tw/˜cjlin/papers/libsvm.ps.gz