Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method

Feng Gao and Haizhou Ai

Computer Science and Technology Department,
Tsinghua University, Beijing 1000084, China
gaofeng03@mails.tsinghua.edu.cn, ahz@mail.tsinghua.edu.cn

Abstract. As we all know, face age estimation task is not only challenging for computer, but even hard for human in some cases, however, coarse age classification such as classifying human face as baby, child, adult or elder people is much easier for human. In this paper, we try to dig out the potential age classification power of computer on faces from consumer images which are taken under various conditions. Gabor feature is extracted and used in LDA classifiers. In order to solve the intrinsic age ambiguity problem, a fuzzy version LDA is introduced through defining age membership functions. Systematic comparative experiment results show that the proposed method with Gabor feature and fuzzy LDA can achieve better age classification precision in consumer images.

Keywords: Age classification, Gabor feature, membership functions, fuzzy LDA.

1 Introduction

As one of the main human facial attributes, aging plays a more complex role than other factors such as human identity, expression, gender or race. The progress of human aging is uncontrollable, with many internal and external influence factors such as one's health state, lifecycle and extreme weather conditions. Besides, because age is a temporal property of people, it's hard to collect the same person's face image across ages. It's also tedious and laborious to label the exact or approximate ages of collected faces. Due to these present difficulties, researches on human age are not as much as that on other face attributes.

However, researches on age progression and estimation have large potential in many applications, e.g. homeland security, parental control, age based Human-Computer interaction, passports renewal and finding missing individuals, and in particular face retrieval over internet or large scale face image database that is our targeted application area. In literature, different aspects of age progression are researched, including building complex models to predicate or simulate one's facial appearance in future [1], estimating or classifying the age of a given face image [2][4][5][6], age progression modeling to alleviate performance drop in face recognition [3][7].

Our work focus on the face age classification problem, the most related work with ours is Fu et al.'s work on age estimation [5] and Yang et al.'s work on age classification [6]. Fu et al. [5] applied linear dimensional reduction algorithm to map human

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 132–141, 2009.

faces from pixel intensity space to a smaller space, which facilitated the following age regression using quadratic function or SVR [4], they achieved an average error of about 5 years in age estimation on a private large database. However, the age estimation variation across ages is large, without any result on coarse age classification task. Yang et al. used LBP feature and AdaBoost algorithm to construct a classifier to classify face as one of the three coarse categories: child, adult and old people. His training set is built from snapshot faces of Asian people taken under constraint illumination condition and without any expression and pose variations, thus the age classifier has limited performance in consumer images.

In our work, we partition the age into four categories, which are baby (0 to 1 approximately), child (2 to 16 approximately), adult (17 to 50 approximately), and old (after 50 approximately). We collected thousands of frontal or near frontal face images as training set from consumer images, there are variations in illumination and expression among those faces. One thing to mention, the age labeled in our collected data set is subjective one, not the objective one due to lack of the exact age information. Then we use Gabor features [9] as face representation and linear discriminant analysis (LDA) [10] to construct the final age classifier, achieving as much as 91% in precision on test set. Besides, to cope with the age ambiguity, we also employed the concept of fuzzy LDA classification by defining fuzzy age membership functions which not only utilize faces with vague ages to enlarge the training set, but also boost the estimation precision to a higher level significantly.

The rest of the paper is organized as follows: section 2 describes the Gabor features used for age classification; in section 3 fuzzy version LDA is presented; section 4 gives systematic comparative experiment results on a large consumer image dataset; and section 5 draws the conclusion.

2 Gabor Features

Gabor features [9] are popular in face representation; its effectiveness has been proved by many researches in fields like face recognition [9]. In face age classification, we also choose to use Gabor features, and in particular we extract Gabor features of 3 scales and 4 orientations that amount to 12 convolved face images of which only magnitude images are used as raw features as shown in Figure1. PCA is used to raw features for dimension reduction.

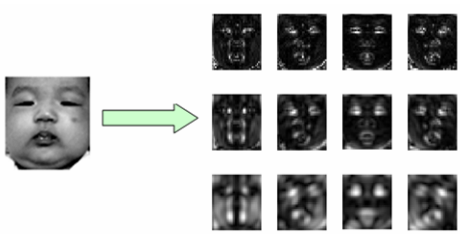

Fig. 1. The Gabor magnitude images of a baby face

Gabor filters are defined as below:

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

where $k_{\mu,\nu} = k_\nu e^{i\phi\mu}$, $k_{\mu,\nu} = k_{\max} / \lambda^\nu$, $\phi_\mu = \pi\mu / 8$.

3 Fuzzy LDA

Linear discriminant analysis (LDA) as a classical dimension reduction method aims to find out optimal project directions to maximize the ratio of the between-class scatter and the within-class scatter. After finding the projected directions, data can be mapped to a low-dimensional subspace, and the nearest class center criteria can be used for classification.

In LDA method, each training sample is assigned to one class label exactly, this is easy in many other classification problems, because these classes have been clearly defined. But in age classification, mapping ages to age groups is somewhat intrinsically ambiguous. For example, it's usually quite easy to judge a face which is 6 years old as child, but how about the one with the label of 14, or what age group is appropriate for a 50 years old man. With labeled faces at hand, we need a way to map an age label to an appropriate age group to build a training set. There are three kinds of way to this goal: 1) Drop some vague ages; 2) Assign every age to only one age group. 3) Assign every age to age groups with the help of fuzzy age membership functions. In fact, the above three can all be expressed through age membership functions based on fuzzy mathematics. With the fuzzy age membership functions introduced, we modified the LDA method to a fuzzy version to suit the age classification problem.

3.1 Age Membership Functions

In order to cope with age ambiguity, we defined an age membership function $\mu_i(x)$ as:

$$0 \le \mu_i(x) \le 1, \quad i \in \{1,2,3,4\}, \quad 0 \le x, \quad \sum_{i=1}^{4} \mu_i(x) = 1$$

It describes to what extent a face with an age label *x* is a member of *i-th* age group. Note that the age ambiguity always happens at two adjacent age group's boundary, each age can belong to at most two age groups, therefore we elaborately designed the 3rd kind of age membership functions , as well as the other two non-fuzzy kind of membership functions as follows and their curves are shown in Figure 2.

(1) The 1st kind of age membership functions when some vague ages are dropped

$$\mu_1(x)=\begin{cases}1 & x=0\\ 0 & else\end{cases};\quad \mu_2(x)=\begin{cases}1 & 3\le x\le 12\\ 0 & else\end{cases};$$

$$\mu_3(x)=\begin{cases}1 & 20\le x\le 40\\ 0 & else\end{cases};\quad \mu_4(x)=\begin{cases}1 & x\ge 60\\ 0 & else\end{cases}$$

(2) The 2nd kind of age membership functions when every age is assigned to exactly one age group

$$\mu_1(x)=\begin{cases}1 & x=0\\ 0 & else\end{cases};\quad \mu_2(x)=\begin{cases}1 & 1\le x\le 17\\ 0 & else\end{cases};$$

$$\mu_3(x)=\begin{cases}1 & 18\le x\le 55\\ 0 & else\end{cases};\quad \mu_4(x)=\begin{cases}1 & x\ge 56\\ 0 & else\end{cases}$$

(3) The 3rd kind of fuzzy age membership functions

$$\mu_1(x)=\begin{cases}1 & x=0\\ 0.8 & x=1\\ 0.5 & x=2\\ 0 & else\end{cases};\quad \mu_2(x)=\begin{cases}0.2 & x=1\\ 0.5 & x=2\\ 1 & 3\le x<12\\ 1-1.5(\frac{x-12}{8})^2 & 12\le x<16\\ 1.5(\frac{x-20}{8})^2 & 16\le x<20\\ 0 & else\end{cases}$$

$$\mu_3(x)=\begin{cases}1.5(\frac{x-12}{8})^2 & 12\le x<16\\ 1-1.5(\frac{x-20}{8})^2 & 16\le x<20\\ 1 & 20\le x<36\\ 1-1.2(\frac{x-36}{24})^2 & 36\le x<48\\ 1.2(\frac{x-60}{24})^2 & 48\le x<60\\ 0 & else\end{cases};\quad \mu_4(x)=\begin{cases}1.2(\frac{x-36}{24})^2 & 36\le x<48\\ 1-1.2(\frac{x-60}{24})^2 & 48\le x<60\\ 1 & x\ge 60\\ 0 & else\end{cases}$$

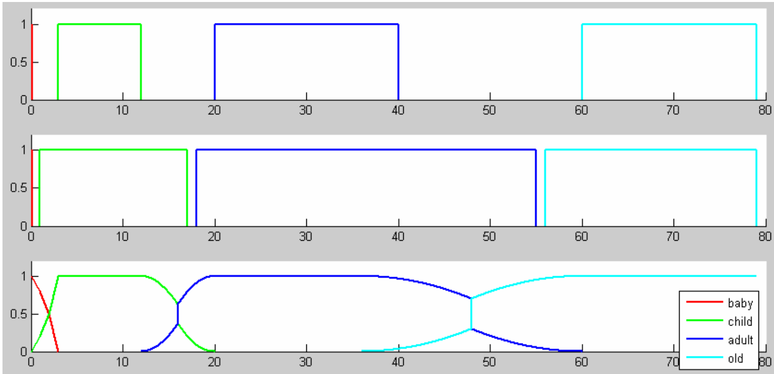

Fig. 2. Three kinds of fuzzy age membership functions: on the top, only non-vague ages are assigned to a certain age group, while in the middle, all ages are belonging to exactly one group, at the bottom is the more natural one being fuzzy because it takes into account of the age ambiguity

3.2 Fuzzy LDA Method

With age membership functions defined, they can be used in LDA by way of using age membership functions as class weighting values as follows:

$$S_B = \sum_{i=1}^{c} N_i^{'}(m_i - m)(m_i - m)^T,\ S_w = \sum_{j=1}^{C}\sum_{i=1}^{N} \mu_j(x_i)(x_i - m_j)(x_i - m_j)^T$$

where $m = \frac{1}{N}\sum_{i=1}^{N} x_i$, $m_i = \sum_{i=1}^{N}\mu_j(x_i)x_i \Big/ \sum_{i=1}^{N}\mu_j(x_i)$, $N_j^{'} = \sum_{i=1}^{N}\mu_j(x_i)$, N is the total number of training instances. As conventional LDA, the optimal projections are defined as below.

$$W_{opt} = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1, w_2, \ldots w_m]$$

where $[w_1, w_2, \ldots w_m]$ is the set of generalized eigenvectors of $S_W^{-1} S_B$.

4 Experimental Results

A face image data set is collected from Internet, and divided into a training dataset and a test dataset. Using face detector and face alignment tool, these faces are automatically cropped and normalized in grey level and geometry as in [6], and each face is manually labeled with an age value estimated by human subjectively. The final training dataset consists of 5408 faces of 64 by 64 in resolution and the labeled ages range from 0 to 79 years old, while the test dataset consists of 57 babies, 350 children, 492 adults and 79 old people that amount to totally 978 photos. For the four-class classification, faces in the training dataset will be assigned to age groups according to their labeled age. Due to every face has an age value labeled, not only it can be assigned a certain age group according to the class membership functions, but also it can provide necessary data for building an age regression model. In our experiments, four aspects are approached: 1. what feature is most effective in age classification? 2. Does Fuzzy LDA help in improving age classification precision? 3. What's the performance of other classification methods? 4. Does age regression help in age classification?

4.1 Comparative Experiment on Different Features

Besides Gabor feature, we also used two other features: pixel intensity and LBP [8]. PCA is used for dimension reduction as preprocessing. We use the 1st kind of age membership functions defined in section 3.1 to label the whole collected face image set with age group attribute, which in fact is a subset of the original dataset containing 644 babies, 1427 children, 1691 adults and 1025 elder people by excluding some vague ages for better separation. The conventional LDA is used to extract a 3D discriminative feature space for each kind of feature. Figure 3 visualized the training data. It is clearly that Gabor feature is more discriminative than the other two. And the quantitative results in Table 1 (using the test set mentioned above) give more support to Gabor feature in its discriminant power and the generalization performance on the age classification problem.

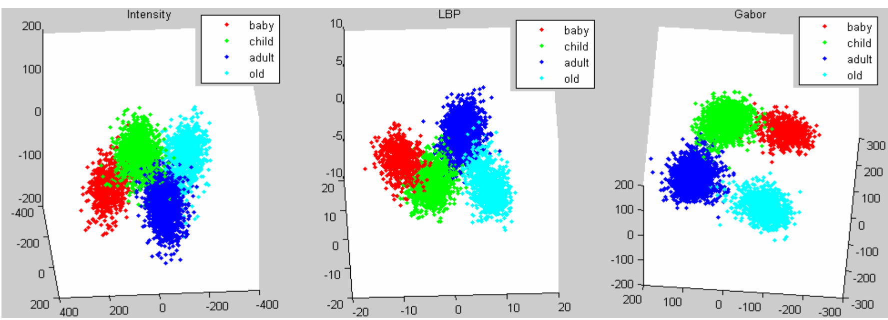

Fig. 3. The training data visualization in discriminative feature space (left for intensity feature, middle for LBP feature, right for Gabor feature)

Table 1. Age classification precision using various features

Dataset	Training set			Test set		
feature	intensity	LBP	Gabor	intensity	LBP	Gabor
Baby	95.34%	97.67%	98.29%	91.07%	96.43%	94.74%
Child	94.81%	97.06%	98.95%	89.14%	79.71%	92.57%
Adult	95.86%	96.87%	98.94%	88.38%	88.18%	91.46%
Old	95.12%	96.49%	99.71%	70.51%	75.64%	78.48%
Total precision	95.32%	96.95%	99.02%	87.39%	84.64%	91.00%

4.2 Comparative Experiments on Fuzzy LDA

In this experiment, we only use the Gabor feature which has been proved to be more effective than other features. We carry on our experiments using three kinds of membership functions defined in section 3.1, notice that the fuzzy LDA method using the 1st kind of age membership functions is equivalent to the LDA method with four explicit classes. Table 2 gives the result. By using the 2nd kind of age membership functions, although the training set is enlarged, the added faces are assigned to exactly one age group, the prediction performance on test set is dropped, especially for the baby class. While using the 3rd kind of age membership functions which are more smooth and natural, the prediction performance get improved to 92.54%, especially with old people class boosting from 78.48% to 88.61% and the children class from 92.57% to 94.57%, and the precision doesn't drop for adult and baby. From this result, it is obvious that the introduction of fuzzy LDA helps to utilize the face data with vague ages and improves the age classification generalization performance. In Figure 5 it shows some demo pictures containing the age classification result using our fuzzy LDA method with Gabor features.

Table 2. Comparative result using different age membership functions

Membership functions	(1)	(2)	(3)
Baby	94.74%	61.40%	94.74%
Child	92.57%	94.57%	94.57%
Adult	91.46%	90.04%	91.46%
Old	78.48%	77.22%	88.61%
Total precision	91.00%	88.96%	92.54%

4.3 Comparative Experiments with SVM, AdaBoost and LDA

Other than LDA method, there are many effective classification methods among which Support Vector Machines (SVM) [13] and Real AdaBoost [12] are two popular representatives. In this experiment, on the same training and test set as in section 4.1

Table 3. Comparative result using SVM, AdaBoost and LDA

Method	SVM		AdaBoost		LDA
Feature	One-vs-All (Gabor)	One-vs-Another (Gabor)	Haar	Gabor	Gabor
Baby	78.95%	85.96%	82.46%	75.44%	94.74%
Child	91.43%	91.71%	80.86%	71.14%	92.57%
Adult	88.41%	88.82%	80.08%	81.71%	91.46%
Old	86.08%	89.87%	73.42%	70.89%	78.48%
Total precision	88.75%	89.78%	79.96%	76.69%	91.00%

we report their results on age classification. For multi-class classification, we use a binary tree structure classifier as in [6] in which each node is trained by Real AdaBoost, and also pair-wise SVM classifiers using polynomial kernel trained with the “One-vs-Another” strategy that vote for the final decision [15] which got the best performance among different kernels including linear and Gaussian and also with the “One-vs-All” strategy [15]. In both methods Gabor features are used, and for AdaBoost we also used Haar [14] feature. Table 3 shows the results, in which we can see the LDA is comparative with the SVM, and it is much better than the AdaBoost in test set. For accuracy and efficiency, the LDA method is preferred in our age classification problem.

4.4 Experiment on Age Regression

Fu et al.’s work [4] on age estimation resulted in about 5 years old of mean absolute error. However, in their paper, the curve of the mean absolute errors in each age is heavily vibrating, with large errors in some ages even more than 10 years old. That means the precision of age estimation by regression has great variation over ages, together with a mean absolute error that is not so small for age classification requirement. Therefore we are not sure whether age regression method could help in age classification.

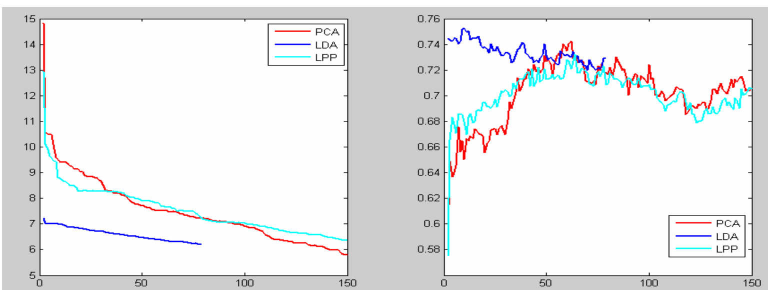

Fig. 4. Left side shows the age error in a 5-fold cross variation set, and right side shows the age classification precision in test set. The x-coordinate of both figures indicates the reduced dimension, and from left to right, the y-coordinate represents mean age error and classification precision respectively.

To be sure, we implemented the age regression method in which Gabor features are used as face representation, and then PCA, LDA and LPP (Locality Preserving Projections [11]) are used for linear dimensional reduction respectively and then Quadratic function [5] is used for age regression. Result in Figure 4 shows that the least age absolute error is 5.8061 and the best age classification precision on the test set is 75.26%, which is not comparable with that using the Fuzzy LDA. The reason is perhaps that variation related to age is too complicated to build a unified age estimation model. With the limited accuracy and vibrated performance, regression approaches for coarse classification task do not make sense.

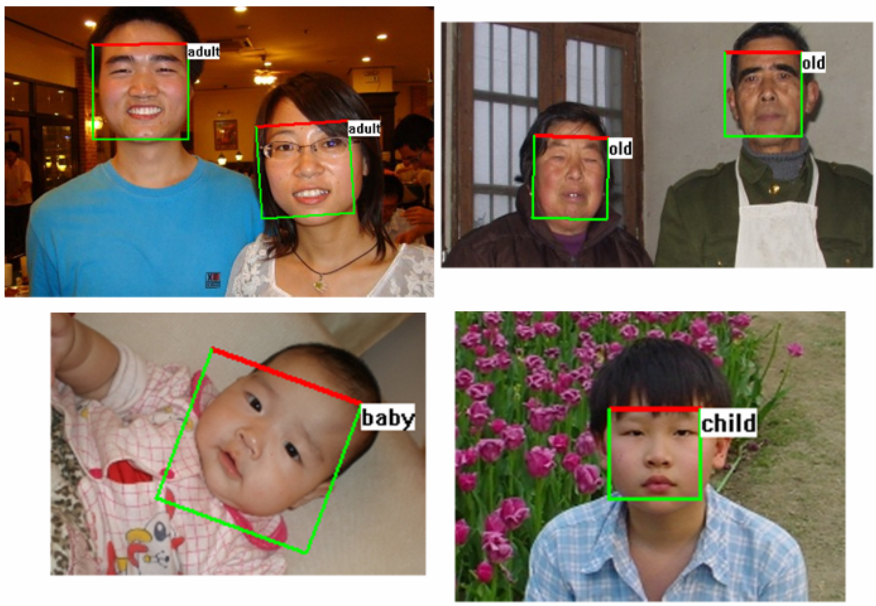

Fig. 5. Some age classification results using Gabor feature and Fuzzy LDA

5 Conclusion

In this paper, by introducing fuzzy membership functions, we proposed a Fuzzy LDA method using Gabor features for coarse age classification. Comparative experiments on different features and different age membership functions show that Gabor feature outperforms other features like pixel intensity and LBP, and the Fuzzy LDA can improve the classification precision even further. Besides, comparative experiments using SVM, AdaBoost and LDA have been done, to show that the LDA method works better for age classification. In addition, we have proved that for age classification it is more effective to use a discriminant classification method rather than to build a unified age regression model.

Acknowledgement

This work is supported by National Science Foundation of China under grant No.60673107, and it is also supported by a grant from HP Corporation.

References

1. Suo, J., Min, F., Zhu, S., Shan, S., Chen, X.: A Multi-Resolution Dynamic Model for Face Aging Simulation. In: CVPR 2007 (2007)
2. Geng, X., Zhou, Z.-H., Smith-Miles, K.: Automatic Age Estimation Based on Facial Aging Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(12) (December 2007)

3. Wang, J., Shang, Y., Su, G., Lin, X.: Age simulation for face recognition. In: ICPR 2006 (2006)
4. Guo, G., Fu, Y., Dyer, C.R., Huang, T.S.: Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression. IEEE Transactions on Image Processing 17(7) (July 2008)
5. Fu, Y., Huang, T.S.: Human Age Estimation With Regression on Discriminative Aging Manifold. IEEE Transactions on Multimedia 10(4) (June 2008)
6. Yang, Z., Ai, H.: Demographic classification with local binary patterns. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 464–473. Springer, Heidelberg (2007)
7. Ling, H., Soatto, S., Ramanathan, N., Jacobs, D.W.: A Study of Face Recognition as People Age. In: ICCV 2007 (2007)
8. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
9. Liu, C., Wechsler, H.: Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. IEEE Transactions on Image Processing 11(4) (April 2002)
10. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1064. Springer, Heidelberg (1996)
11. He, X., Niyogi, P.: Locality Preserving Projections. In: NIPS 2004 (2004)
12. Schapire, S.: Improved Boosting Algorithms Using Confidence-rated Predictions. Machine Learning 37(3), 297–336 (1999)
13. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, Heidelberg (1995)
14. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: CVPR 2001 (2001)
15. Zhuang, L., Ai, H., Xu, G.: Training Support Vector Machines for Video Based Face Recognition. In: International Conf. on Image and Graphics. SPIE, vol. 4875, pp. 737–743 (2002)