Facial Image Retrieval Based on Demographic Classification

Bo WU, Haizhou AI and Chang HUANG
Dept. of Computer Science and Technology, Tsinghua University,
State Key Laboratory of Intelligent Technology and Systems，Beijing 100084, PR China
E-mail: ahz@mail.tsinghua.edu.cn

Abstract

In this paper we propose a novel method for demographic classification and present an image retrieval system that can retrieve facial images by demographic information that includes gender, age and ethnicity. The demographic information is extracted from human faces by demographic classifiers that are learned from boosting Haar feature based Look-Up-Table type weak classifiers. The image retrieval system consists of three modules, face detection, facial feature landmark extraction and demographic classification. Experiment results are reported to show its potential in the management of a large facial image database for online retrieval applications.

1. Introduction

Facial image retrieval retrieves images based on information extracted from human faces. It is a specific problem of content based image retrieval and has great potential in various applications, such as HCI, digital video processing and visual surveillance. Due to the rising popularity of digital cameras and digital albums, retrieving images of human faces becomes an interesting problem. In the literature, Gudivada et al. [1] proposed a framework for image retrieval systems and implemented a feature based face retrieval system; Satoh [2] built a face retrieval and recognition system called "Name It" for video content analysis based on an eigen-face method for face recognition. Eickeler [3] used a pseudo 2D Hidden Markov Model to retrieve faces from a face database.

Most of the current face retrieval systems are based on the face recognition technique, i.e. retrieval by identity. Besides identity, the face contains a lot of other useful category information, such as gender, age, ethnicity and even expression, etc. It would be helpful to use clues other than identity to retrieve facial images. Demographic information including gender, age and ethnicity that are categories of higher level knowledge than identity could be well suited for retrieval uses. The process of extracting demographic information from a facial image is called demographic classification [4], see Fig.1. Previous works on demographic classification are mainly on gender classification, among which the main methods used include PCA [5], ANN [6], SVMs [7] and Adaboost [4]. For age and ethnicity classification only a few systems have been reported [4][8][9].

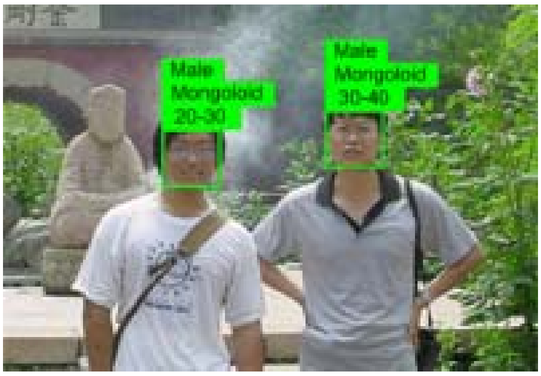


Figure 1. Demographic classification (above the face block the estimations of gender, ethnicity and age are labeled.)

The face retrieval system presented in this paper consists of three modules, face detection, facial feature landmark extraction and demographic classification. The rest of this paper is organized as follows: Section 2 introduces the learning algorithm we used, Adaboost.MH; Section 3 describes the Haar feature based LUT weak classifiers; Section 4 gives the experiment results of demographic classification; Section 5 presents our face retrieval system; Section 6 gives the conclusion and discussion.

2. The Adaboost.MH Algorithm

Adaboost [11] is a leaning algorithm that selects a small number of weak classifiers from a large weak classifier pool or hypothesis space to construct a strong classifier. The two-class Adaboost algorithm has been successfully used in face detection [10] and demographic classification [4]. However, many problems such as ethnicity classification are multi-class in nature. Fortunately, based on a measure of Hamming loss, Schapire and Singer have extended Adaboost to a multi-class multi-label version, called the Adaboost.MH algorithm [12]. Denote the sample space by $\mathcal{X}$ and the label set by $\mathcal{Y}$. A sample of a multi-class multi-label problem is a pair $(\mathbf{x}, Y)$, where $\mathbf{x} \in \mathcal{X}, Y \subseteq \mathcal{Y}$. For $Y \subseteq \mathcal{Y}$, define $Y[l]$ for $l \in \mathcal{Y}$ as

$$Y[l] = \begin{cases} 1 & \text{if } l \in Y \\ -1 & \text{if } l \notin Y \end{cases} \tag{1}$$

Fig. 2 lists the details of the Adaboost.MH algorithm.

Given: the sample set $S = \{(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_m, Y_m)\}$ and the size of the final strong classifier T.

Initialize: $D_1(i,l) = 1/(mk)$, i=1..m, l=1..k, where $k = |\mathcal{Y}|$.

For t = 1,…,T

- Under the distribution D_t, select a weak classifier $h_t : \mathcal{X} \times \mathcal{Y} \to [-1,1]$ from the weak classifier pool to maximize the absolute value of

$$r_t = \sum_{i,l} D_t(i,l) Y_i(l) h_t(\mathbf{x}_i, l). \tag{2}$$

- Let: $$\alpha_t = \frac{1}{2} \ln\left(\frac{1+r_t}{1-r_t}\right). \tag{3}$$

- Update the distribution:

$$D_{t+1}(i,l) = \frac{D_t(i,l)\exp\left(-\alpha_t Y_i[l] h_t(\mathbf{x}_i, l)\right)}{Z_t}, \tag{4}$$

where Z_t is a normalization factor so that D_{t+1} is a p.d.f.

Output the final strong classifier:

$$H(\mathbf{x}, l) = \operatorname{sign}\left(\sum_t \alpha_t h_t(\mathbf{x}, l)\right) \tag{5}$$

Figure 2. The Adaboost.MH algorithm

From Eq.5, it can be seen a sample $\mathbf{x}$ can belong to several classes. To get a multi-class single-label classifier, the Eq.5 is adjusted to the following form:

$$H(\mathbf{x}) = \arg\max_l \left(\sum_t \alpha_t h_t(\mathbf{x}, l)\right), \tag{6}$$

and furthermore a confidence of H is defined as:

$$Conf_H(\mathbf{x}) = \left|\frac{\sum_t \alpha_t h_t(\mathbf{x}, H(\mathbf{x}))}{\sum_t \alpha_t}\right|. \tag{7}$$

3. Haar Feature Based LUT Weak Classifier

In order to use the Adaboost.MH algorithm, a weak classifier pool should be configured. We construct our weak classifier based on the Haar feature proposed by Viola and Jones [10]. For each Haar feature, one weak classifier is configured. In [10], a threshold-type weak classifier, whose output is Boolean value, is used, that is, $h(\mathbf{x}) = \operatorname{sign}\left[f_{Haar}(\mathbf{x}) - b\right]$, where f_{Haar} is the Haar feature and b is a threshold, see Fig.3.a. The main disadvantage of this threshold model is that it is too simple to fit complex distributions, such as a multi-Gaussian. For a multi-class problem, $\varpi_1, \ldots, \varpi_k$, we propose a real-valued 2D LUT-type weak classifier, see Fig.3.b. Assuming f_{Haar} has been normalized to [0, 1], the range is divided into n sub-ranges:

$$bin_j = [(j\text{-}1)/n,\ j/n),\ j = 1,\ldots,n \tag{7}$$

The weak classifier $h(\mathbf{x}, l)$ is defined as

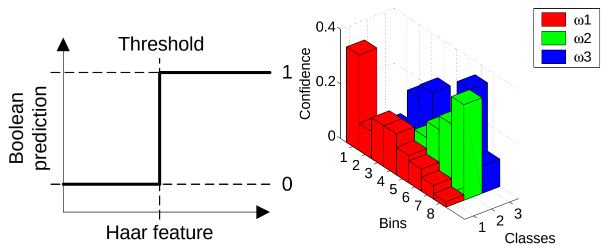
$$\text{If } f_{Haar}(\mathbf{x}) \in bin_j \text{ then } h(\mathbf{x}, l) = 2P_l^{(j)} - 1, \tag{8}$$

where $P_l^{(j)} = P(\mathbf{x} \in \varpi_l \mid f_{Haar}(\mathbf{x}) \in bin_j)$. Given the characteristic function

$$B_n^{j,l}(u, y) = \begin{cases} 1 & u \in \left[(j-1)/n,\ j/n\right) \wedge y = l \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

the LUT weak classifier can be formally expressed as:

$$h_{LUT}(\mathbf{x}, y) = \sum_{j=1}^{n} \sum_{l=1}^{k} \left(2P_l^{(j)} - 1\right) B_n^{j,l}\left(f_{Haar}(\mathbf{x}), y\right). \tag{10}$$



(a) Threshold (b) LUT

Figure 3. Threshold vs. LUT (8-bin 2D LUT for 3-class problem)

In practice, there are usually millions of possible Haar features, so the weak classifier pool can be very large. If exhaustive search is used in each boosting round to find the global maximum, it will be very time consuming. Therefore some optimization technique such as simulated annealing is becoming necessary.

4. Experiment Results

As mentioned above, the demographic classification has three sub-problems, gender, ethnicity and age. The gender classification is a standard two-class problem. In this paper, the ethnicity classification is defined as a three-class problem, Mongoloid (yellow), Caucasian (white) and African (black). Due to the ground truth data of age classification is difficult to obtain, at present the age classification is set as a two-class problem, kid vs. adult. From World Wide Web we have collected a large face dataset that contains 5,597 females and 4,928 males, 2,411 Mongoloids, 2,306 Caucasians and 1,771 Africans, and 1,303 kids and 1,509 adults. All the face samples are normalized to 36×36-pixel patches, see Fig.4. We do not use any color information for generality. For testing, 5-fold cross validation is adopted and the RBF-kernel SVMs method is used for a comparison. For the gender and age classifiers, we set T = 400 in the Adaboost.MH, i.e. 400 weak classifiers will be selected to construct one strong classifier; for the ethnicity classifier T = 600. Experiment results are listed in Table 1. It can be seen that the correctness of our method is slightly better than that of the SVMs method. However in average our method takes only 0.3ms to process a sample on an Athlon 1.4G Hz PC which is faster than the SVMs method by a factor of 400.

Figure 4. Standard face sample: a) female, b) male, c) Mongoloid, d) Caucasian, e) African, f) kid, g) adult.

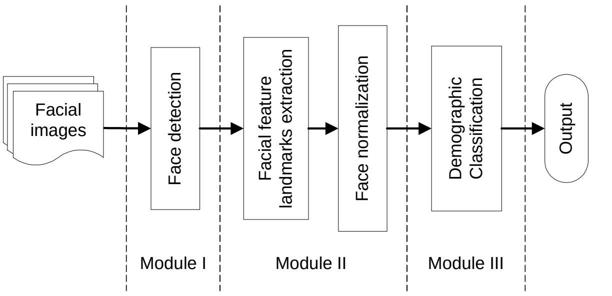
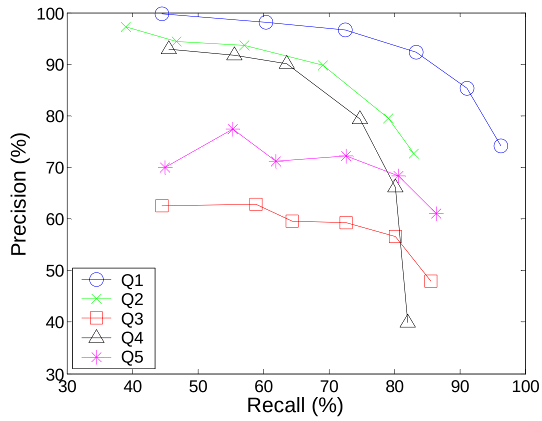


Figure 5. An automatic demographic classification system.

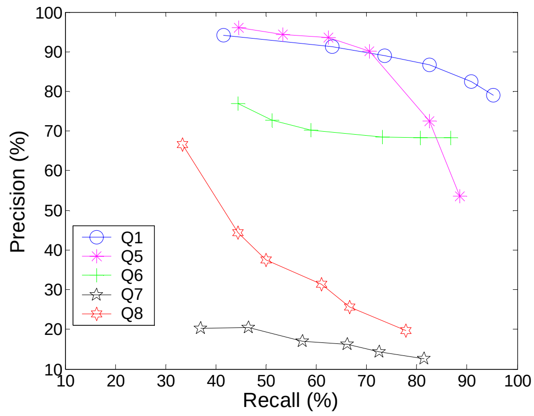
5. Facial Image Retrieval System

To retrieve faces by demographic classification, we need to extract the gender, ethnicity and age information from facial images automatically. The flowchart of this procedure is shown in Fig.5. The system contains three modules, face detection, facial feature landmark extraction and demographic classification. The face detection module is also based on a boosted cascade of LUT weak classifiers [13]. As for facial feature landmark extraction used in face sample normalization, we train a simplified Direct Appearance Model (SDAM) [14] to extract three facial landmarks, the pair of eyes and the mouth center that are used as referenced landmarks for warping to a normalized shape. For demographic classification, we use the demographic classifiers described in the Section 4.

According to Eq.7, for each prediction the demographic classifier can give a real-valued confidence between [0, 1]. This confidence is very useful for face retrieval. A single query of our retrieval system has two parts, a class label and a threshold. For example the pair ('Female', 0.3) is a single query indicating that we want to find all females with prediction confidences over 0.3. A valid query consists of a series of such pairs, e.g. {("African", 0.1), ("Male", 0.2)}. All samples satisfying the query will be ranked according to their confidences. Our method can achieve a processing speed of about 40ms on a 320×240 image on an Athlon 1.4GHz PC.



(a) On FERET test set



(b) On WWW test set

Figure 6. The precision-recall curves for some queries: Q1: Female, Q2: African, Q3: African Male, Q4: Caucasian Female, Q5: Mongoloid Male, Q6: Female Adult, Q7: Mongoloid Kid, Q8: African Female Adult.

We have used two data sets to test our system, one from FERET [15] containing 3,105 faces and the other from WWW containing 1,753 faces. Fig.6 gives the precision-recall curves for some queries and Fig.7 shows the first six faces found. It can be seen that the system has better performance on the FERET set than on the WWW set. One possible reason for this is that all the images in the FERET set are snapshots captured under controlled environments while most of the images in the WWW set are real-life photos or posters that have relatively large variance on pose, expression and illumination, etc.

6. Conclusion

In this paper we have proposed a LUT weak classifier based boosting method for face retrieval by demographic classification. We have developed a Haar feature based 2D LUT-type weak classifier for multi-class problems and use a variation of boosting algorithm for multi-class multi-label problems, Adaboost.MH, to learn the demographic classifiers. And a prototype of automatic demographic face retrieval system is presented and experiment results show its potentials in the management of a large facial image database for online retrieval applications.

7. Acknowledgements

This work was supported by National Science Foundation of China under grant No.60332010.

8. References

[1] V. N. Gudivada and V. V. Raghavan. "Modeling and Retrieving Images by Content". Info. Proc. and Manag., 33(4):427--452, 1997.

[2] S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming and Detecting Faces in News Videos", IEEE MultiMedia, 6(1):22-35, 1999.

[3] S. Eickeler, "Face Database Retrieval Using Pseudo 2D Hidden Markov Models", in IEEE conf. on FGR2002.

[4] G. Shakhnarovich, P. Viola and B. Moghaddam, "A Unified Learning Framework for Real Time Face Detection and Classification", in FGR2002.

[5] Alice J.O'Toole et al. "The Perception of Face Gender: The Role of Stimulus Structure in Recognition and Classification". Memory and Cognition, Vol. 26, pp. 146-160, 1997.

[6] A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces". In Advances in Neural Information Processing Systems, pp. 572–577, 1991.

[7] B. Moghaddam and M.H. Yang, "Gender Classification with Support Vector Machines", IEEE Trans. on PAMI, Vol. 24, No. 5, pp. 707-711, 2002.

[8] S. Gutta, H. Wechsler, and P. J. Phillips, "Gender and ethnic classification", IEEE conf. on FG1998.

[9] Jun-ichiro Hayashi, et al., "Age and Gender Estimation based on Wrinkle Texture and Color of Facial Images", In ICPR2002.

[10] Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", IEEE conf. on CVPR2001.

[11] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm", In Proceedings of the 13-th Intern. Conf. on Machine Learning, pp. 148-156, Morgan Kaufmann, 1996.

[12] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions", Machine Learning, 37:297-336, 1999.

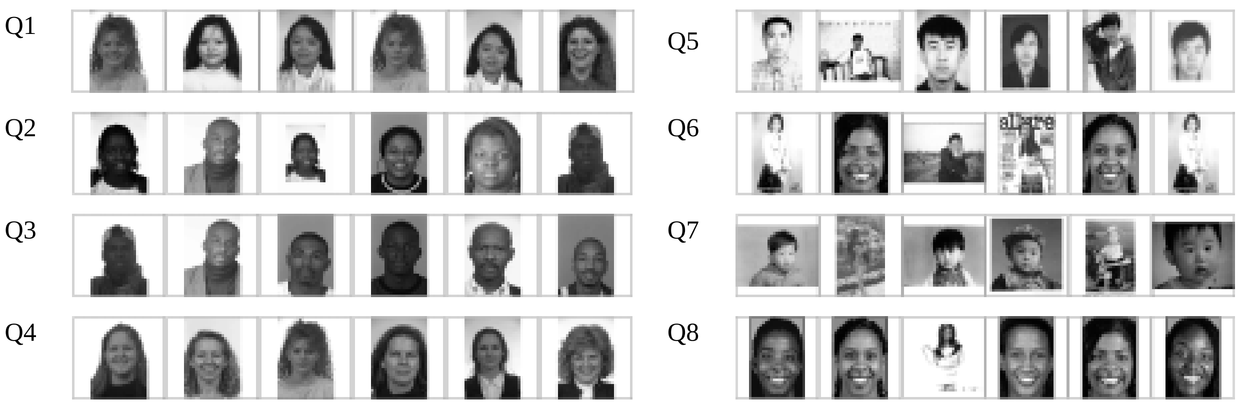
[13] B. Wu, H. Z. Ai, C. Huang, S. H. Lao, "Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost", In Proc. the 6th IEEE Conf. on Automatic Face and Gesture Recognition (FG 2004), Seoul, Korea.

[14] T. Wang, H. Z. Ai, G. F. Huang, "A Two-Stage Approach to Automatic Face Alignment", in Proc. of SPIE Vol. 5286 Third International Symposium on Multispectral Image Processing and Pattern Recognition, pp 558-563, 2003

[15] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The *FERET* database and evaluation procedure for face recognition algorithms", Image and Vision Computing J, Vol. 16, No. 5, pp 295-306, 1998.

Table 1. Results of demographic classification

Problem / Method	Gender			Ethnicity				Age		
	Female	Male	Overall	Mongoloid	Caucasian	African	Overall	Kid	Adult	Overall
Ours	92.3%	90.2%	91.4%	95.0%	96.1%	93.5%	95.0%	95.8%	95.9%	95.8%
SVMs	91.6%	90.9%	91.0%	95.6%	95.4%	92.7%	94.7%	93.7%	94.8%	94.3%



(a) On FERET test set

(b) On WWW test set

Figure 7. First six retrieval results: Q1: Female, Q2: African, Q3: African Male, Q4: Caucasian Female, Q5: Mongoloid Male, Q6: Female Adult, Q7: Mongoloid Kid, Q8: African Female Adult.