The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

A Real-time Face Detection and Recognition System for Mobile Robot in the Complex Background

Song Chen[1,2], Tao Zhang[1,2], Chengpu Zhang[1,2] and Yu Cheng[3]

[1]*Department of Automation, Tsinghua University, Beijing 100084, China*
(Tel: 86-10-6279-7710; Fax: 86-10-6278-6911)
(E-mail: chen-s03@mails.tsinghua.edu.cn)
[2]*Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China*
[3]*Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China*

Abstract: The research presented in the paper focuses on a real-time face detection and recognition system applied to biped robot in the complex background. In the visual system, a multi-information method consisted of Adaboost algorithm and color information for face detection part is proposed and Embedded Hidden Markov Model (EHMM) is employed to recognize the detected faces. The system introduced in the paper improves the processing speed of detecting and recognizing faces in a frame with a suitable accuracy by integrating three rapid algorithms.

Keywords: face recognition, biped robot, complex background

I. INTRODUCTION

In the recent decades, the scope of applying mobile robots has been broadening with the development of robot technologies. The mobile robots can not only get used to the variety of terrestrial environments, but also are required to have more effective capability of interaction with users, for example, following the people's actions or communicating with people [1], which makes the ability of acquiring information by vision be more and more important to robots. Department of Automation, Tsinghua University has developed a biped robot used to research and integrate different robot technologies. The face recognition system introduced in the paper is developed for the robot to make it detect and recognize faces in the complex background.

The character of the system is real-time processing and adapting to uncontrolled indoor environments. For the purpose, a multi-information method consisted of Adaboost algorithm and color information, which are the fast face detection methods at present, for face detection part is proposed. In addition, Embedded Hidden Markov Model (EHMM) is presented using 2-DCT feature vector as the observation vector to recognize the detected faces. By setting the factors of EHMM, the processing speed and recognition rate of the system can be changed. Fig.1 illustrates the main procedure of the face recognition system.

In this paper, Section 2 presents the face detection algorithm, and EHMM algorithm for face recognition is explained in Section 3. In Section 4 the testing results of the whole system are explained in the uncontrolled indoor conditions. Finally, conclusions and future work are drawn in Section 5.

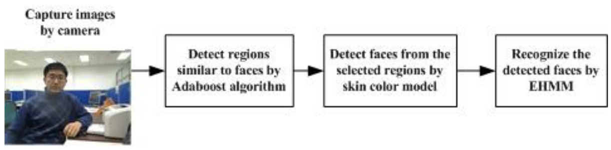

Fig.1. Flow of face detection and recognition

II. FACE DETECTION

The face detection section is integrated by Adaboost algorithm and skin color model method. Both approaches can process the images and video extremely rapidly.

1. Adaboost Algorithm

Adaboost algorithm was first proposed by Freund Y. and Schapire R.E. in 1996, which was an iterative and very important machine learning method [2]. On the basis, Viola, P. and Jones, M. proposed an improved machine learning approach, which could process the images extremely efficiently and achieve high detection rates, for visual objects detection [3][4].

Adaboost detector employs the extended set of Haar-like features which is proposed by Lienhart et al.

[5][6] to classify. Each feature is consisted of 2~3 rectangles to detect edge features, center-surround features and line features. Fig.2 describes the set of features. In order to compute the features rapidly, Viola P. introduced the integral image representation for images. The integral image at location (x, y) means the sum of the pixels above and to the left of (x, y), inclusive:

$$ii(x, y) = \sum_{x' \le x, y' \le y} i(x', y') \qquad (1)$$

where $ii(x, y)$ is the integral image and $i(x, y)$ is the original image.

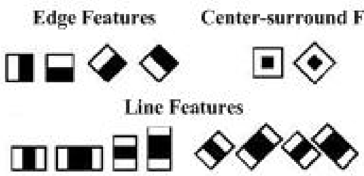

Fig.2. Haar-like features

With the feature set and a training set of positive and negative images, Adaboost algorithm is used to train strong classifiers and a cascade of classifiers is constructed to achieve increased detection performance while reducing computation time. The main procedure is below:

(1) Given example images $(x_1, y_2), \cdots, (x_n, y_n)$, and $y_i = 1, 0$ for negative and positive example respectively.

(2) Initialize weights $w_{1,i} = 1/2m, 1/2l$ for $y_i = 1, 0$ respectively, where *m* and *l* are number of negatives and positives respectively.

(3) For $t = 1, \cdots, T$:

a) Normalize the weights $w_{t,i} \leftarrow w_{t,i} / \sum_{j=1}^{n} w_{t,i}$, so that w_t is a probability distribution.

b) For each feature *j* train a classifier h_j which is restricted to using a single feature. The error is $\zeta_j = \sum_i w_i \left| h_j(x_i) - y_i \right|$.

c) Choose the classifier h_t with the lowest error.

d) Update the weights: $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$, where $e_i = 0$ if example x_i is classified correctly, otherwise $e_i = 1, \beta_t = \zeta_t / (1 - \zeta_t)$.

(4) The final strong classifier is:

$$h(x) = \begin{cases} 1 \sum_{t=1}^{T} \alpha_t h_t(x) \ge (\sum_{t=1}^{T} \alpha_t)/2 \\ 0 \; otherwise \end{cases}, where \; \alpha_t = \log(1/\beta_t)$$

(5) Constructed the cascade of strong classifiers.

2. Skin Color Model

Human skin color has been used and proven to be an effective feature in face detection [7][8]. Several color spaces have been utilized to label pixels as skin including RGB, normalized RGB, HSV, YCrCb, YIQ, CIE XYZ and CIE LUV. The simplest and most effective model is YCrCb model, in which Y means luminance or gray value and Cr, Cb values are defined as skin tone pixels. YCrCb model is influenced little by illumination variations and the distribution of skin color pixels measured by Cr, Cb is very compact. In Fig.3 is shown the distribution according to the statistics of 200 images searched in the Internet.

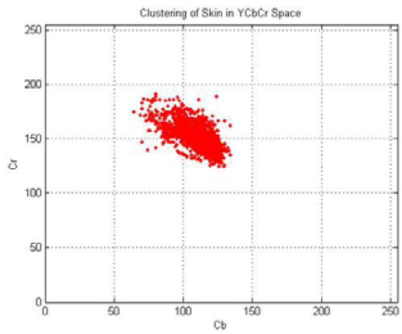

Fig.3. Distribution of skin color in YCrCb model

There are two methods to classify pixels. One is to construct the Gaussian Density Model based on the statistical data, and the other is to classify a pixel simply to have skin tone if its values (Cr, Cb) fall within a range, which is $133 < Cr < 173, 77 < Cb < 127$ concluded by experiments. Considering the detection rates and speed, the second method is employed in the paper. The skin color model is used to verify the regions detected by Adaboost algorithm. If the ratio of skin color pixels and total pixels is larger than the threshold, which is set as 0.355 got by experiments, the regions are considered as faces finally.

3. Face Detection Database

GTAV face database is used to test the effectiveness of face detection section in the paper. GTAV database includes a total of 44 persons with 27 pictures per person which correspond to different pose views (0º, ±30º, ±45º, ±60º, ±90º) under three different illuminations (environment or natural light, strong light source from an angle of 45º, and finally an almost frontal mid-strong light source). Furthermore, at least 10 more additional frontal view pictures are included with different occlusions and facial expression variations.

The resolution of the images is 240*320 with BMP format.

40 persons with 15 images per person with different illuminations and facial expressions are selected to detect, and the people's pose views are limited to 0º, ±30º. Two-thirds of the faces are used to train the cascade of classifiers. In Fig.4 is shown the examples of GTAV and Fig.5 describes the relevant detection results. The detector runs at 4.5 images per second and the detection rate is 97.5%.

Fig.4. Examples of GTAV

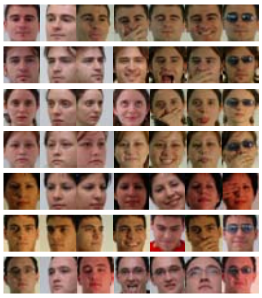

Fig.5. Detection results

III. FACE RECOGNITION

In the face recognition section, Embedded Hidden Markov Model abbreviated as EHMM is employed to recognize the faces detected. Lack of space, the section will just introduce the basic principle and program flow and references [9][10] are given a detailed mathematical derivation of HMM and EHMM.

Hidden Markov Model (HMM) is a set of statistical models used to characterize the statistical properties of a signal. HMM consists of five elements: (1) *N*, the number of states in the model; (2) *M*, the number of different observation symbols; (3) *A*, the state transition probability matrix; (4) *B*, the observation symbol probability matrix; (5) Π, the initial state distribution. With the factors, a HMM is defined as

$$\lambda = (A, B, \Pi) \quad (2)$$

For faces of people, the significant facial regions (hair, forehead, eyes, nose, mouth and chin) come in a natural order from top to bottom regardless of the scale, illumination, rotation of faces. Each of these facial regions is assigned to a state in a left to right 1-demension continuous HMM. Then the observation vectors consist of a set of 2D DCT coefficients that are extracted from each region.

Thus, the HMM $\lambda = (A, B, \Pi)$ is initialized. The training data is uniformly segmented from top to bottom in $N = 5$ *or* 6 states and the observation vectors associated with each state are used to obtain initial estimates of the observation probability matrix B. The initial values for A and Π are set given the left to right structure of the face model.

EHMM is similar to HMM, but consists of s set of super states. Super states divide face into different regions which are hair, forehead, eyes, nose and mouth, from top to bottom. Each super state also consists of some states named embedded states. The faces are divided into some blocks both vertically and horizontally. 2D DCT features are extracted from the blocks as the observation vectors. EHMM utilizes more information of faces than HMM, so the recognition rate of EHMM is higher. In Fig.6 is described the face structure of EHMM while the program flow is shown in Fig.7.

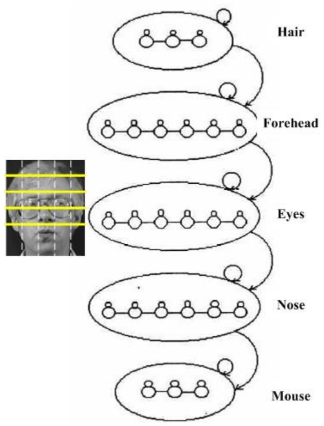

Fig.6. Face structure of EHMM

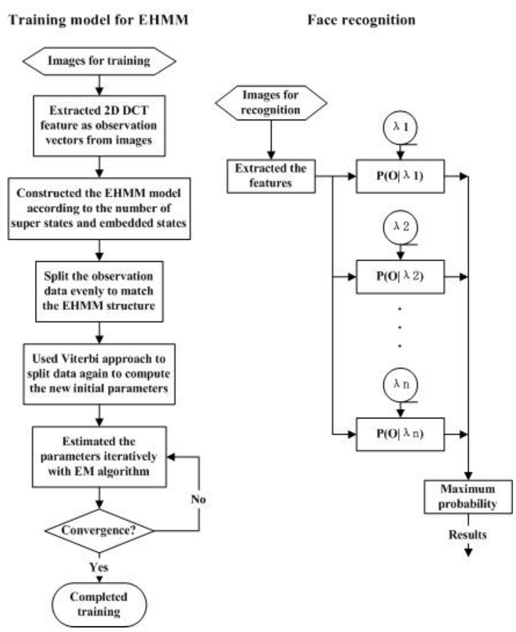

Fig.7. Program flow

ORL (Olivetti Research Ltd.) face database is employed to test the recognition section. ORL database consists of 40 persons with 10 pictures per person. Six pictures per person are used to train the EHMM while the other four images per person are recognized. The recognition results with parameters are shown in table 1.

Table 1. Recognition Results

	Super states number	Embedded states number	Recognition rate	Recognition speed
1	5	3,6,6,6,3	100%	4.1s/image
2	5	1,1,1,1,1	92%	1.3s/image

IV. EXPERIMENTS

The whole system is programmed based on OpenCV under the environment of VC 6.0. The CPU is Pentium Ⅳ with 3GHz frequency and 1.46G memory. The face detection and recognition results are also obtained under the computer's condition. Face database is established by taking photographs of 12 graduate students in the laboratory with 13 pictures per person, of which 8 pictures are used to construct the EHMM. The whole processing of detecting and recognizing a frame, which is 320×240, spends 1.4 seconds with the rapid recognition parameters and 4.2 seconds with the slow recognition parameters. In Fig.8 is described the experiment results, and the frame without yellow words means failure to recognize.

Fig.8. Experiment results

V. CONCLUSIONS

The paper introduces a face detection and recognition system developed for mobile robot to make it have the capability of interaction with people in the uncontrolled indoor environments. The face detection part consists of Adaboost algorithm and skin color model approach to detect faces in the complex background rapidly and efficiently. EHMM algorithm is employed to recognize the detected faces with high recognition rate. The integrated system spends 1.4s processing a frame. Future work will be directed on the improvement of EHMM to increase the recognition speed, and the system should be integrated with the mobile robot. The function of preventing the jitter of video caused by robot's moving also need to be developed.

REFERENCES

[1] E. GARCIA, M.A. JIMENEZ, P. GONZALEZ DE S ANTOS. The Evolution of Robotics Research. IEEE Robotics & Automation Magazine, 2007.3: 2-15.
[2] Freund Y., Schapire R.E. Experiments with a new boosting algorithm. Proceeding of the 13th International Conference on Machine Learning, 1996: 148-156.
[3] Viola, P., Jones, M. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, 2001: 511-518.
[4] Viola, P., Jones, M. Robust real-time face detection. ICCV, 2001.
[5] Lienhart, R., Maydt, J. An extended set of Haar-like features for rapid object detection. ICIP, 2002:900-903.
[6] Lienhart R., Kuranov A. Pisarevsky V, Empirical analysis of detection cascades of boosted classifiers for rapid object. Proceedings of the 25th Pattern Recognition Symposium, 2003: 297 -304.
[7] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. Detecting Faces in Images: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, January 2002: 34-58.
[8] Jean-Christophe Terrillon, Mahdad N. Shirazi, Hideo Fukamachi, et al. Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images. Automatic Face and Gesture Recognition, 2000.
[9] A. Nefian and M. H. H. III. Hidden Markov Models for face recognition. In ICASSP, Vol.5, 1998: 2721–2724.
[10] H. Othman and T. Aboulnasr. Low complexity 2-D Hidden Markov Model for face recognition. In IEEE Int. Symp. on Circuits and Systems, Vol.5, 2000: 33–36.