Robust Realtime Face Recognition And Tracking System

Kai Chen,Le Jun Zhao
East China University of Science and Technology
Email:nasa85220@hotmail.com

Abstract

There's some very important meaning in the study of realtime face recognition and tracking system for the video monitoring and artifical vision. The current method is still very susceptible to the illumination condition, non-real time and very common to fail to track the target face especially when partly covered or moving fast. In this paper, we propose to use Boosted Cascade combined with skin model for face detection and then in order to recognize the candidate faces ,they will be analyzed by the hybrid Wavelet, PCA(principle component analysis) and SVM(support vector machine) method. After that, Meanshift and Kalman filter will be invoked to track the face. The experimental results show that the algorithm has quite good performance in terms of real-time and accuracy.

Keywords;*PCA; meanshift; Kalman filter; svm; wavelet; Realtime face detection; Realtime face tracking; face recognition*

I. INTRODUCTION

Over the last ten years, face recognition and object tracking technology have become a central problem in neural network as well as statistics and signal processing. It's a relatively difficult synthetic problem to implement the face recognition and tracking together, since it requires the synergetic effort of machine learning, pattern recognition, machine vision and image processing. Face is one of the most popular biological features in the current research, which have profound meaning in research and a great potential in application such as safety supervising, video meeting and human-computer interaction. The system should be able to detect, recognize and track the face in realtime. The current methods has several problems like false retrival of faces, the infulence of image noise, the low accurate rate of face recognition, the lack of real time and lose track of the target face. In this paper, we propose an adaptive threshold face detection method which combines the Boosted Cascade with the skin model. After that, face candidate region will be analyzed by the algorithm based on the combination of discrete digital wavelet, principle component analyze and SVM. According to the similarity of the candidate face, the target will be tracked by the algorithm based on the integration of meanshift and Kalman filter. The result of the experiment indicates that the proposed method produces a significant improvement. The rest of the paper is organized as follows. Section 2 provides the detail in boosted cascade with adaptive threshold and skin model for face detection. Section 3 desribes the algorithm composed of DWT,PCA and SVM for face recognition. Section 4 introduces the meanshift and kalman filter for tracking face. The Experimental results are presented in Section 5. Setcion 6 gives the conclusion.

II. BOOSTED CASCADE WITH ADAPTIVE THRESHOLD AND SKIN MODEL FOR FACE DETECTION

The Boosted Cascade still have the limitation of false face retrival, in spite of the fact that it can achieve high performance in face detection. In order to address this problem, We propose a method in which the image is processed by skin model to get the face color region, then the face detecting window will skip the region where the rate of skin color is less than 10 percent, when it is scaned across the image at multiple scales and location. To achieve higher accurate rate, the threshold of boosted cascade is inversely propotional to the skin color rate

so the threshold will be adaptive based on the rate. Although a little bit more expensive in computational term, the false positive rate and residual error rate can be cut down dramatically.

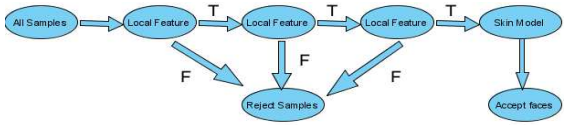

Figure 1. Face Detection FrameWork

Boosted Cascade is a learning-based approach and composed of serveral weak classifiers which can discriminate face and non-face on the basis of Haar-like filters [6] (like figure 2). These filters consist of two or three rectangles. To compute the feature value, the sum of all the pixels values in grey region are substracted from the sum of all the pixels values in white region with the integral image, So the faces can be presented by these rectangle filters in different scale. In this case, the power of classification system can be boosted from several weak classifiers based on simple, local ,Haar-like features.

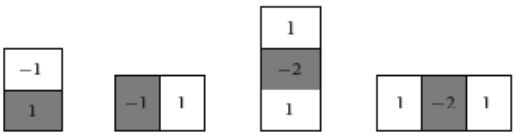

Figure 2. set of Haar-like feature

The Boosted Cascade can achieve relative high perfomance in real-time and accuracy. However, non-face can not be eliminated completely. One way to get rid of non-face region is via the use of the skin model which are quite effective at discriminating skin and non-skin region[7], so the non-face region with the low rate of the color within the range of skin color can be eliminated. To get the skin color region, the RGB space are transformed into Ycbcr and then Y is removed to reduce the influence of illuminance, in the sense that the skin value is relatively centralized in two dimensional space. By solving the distance between the pixel value and the center of gaussian of skin color space , the similarity of pixel value can be computed as

$$P(r,b) = \exp[-0.5(x-m)^T C^{-1}(x-m)], \qquad (1)$$

where m is the mean value, and C is the variance of the skin color space. After that, the threshold is defined by OTSU to get the binary image.

Another obvious defect of boosted cascade is that the threshold in the strong classifier is fixed so that it's very hard to attain an high correct detection rate while obtaining very low false positive rate. To this end, the adaptive threshold of strong classifier is set based on the rate of skin color within the detecting window.

Figure 3. Left: original image. Middle:image based on similarity.Right:binaryimage

By combining the skin model with Boosted Cascade, there's a great improvement in the correct detection rate. In the next section, the candidate face will be analyzed by the hybrid algorithm with wavelet, PCA and SVM.

III. RECOGNIZING FACE WITH WAVELET, PCA AND SVM

In most cases, Face recognition is a two-step process of subspace projecton followed by classification. In this paper, we enhance the face matching technique, for example by pre-processing the image with wavelet and substitue the nearest neighbor classifier with SVM. As the computational intensity of PCA increase sharply with the input size , it will cost too much computation in the stage of training and recognition,especially when the size of face image is quite large. As a result, it won't be fast enough to analyze the face image in realtime. In order to address this problem, we propose to utilize 2 level discrete wavelet transform to obtain the subband representation of the face data by processing face image with

$$\begin{cases} w_{2n}(t) = \sqrt{2}\sum_{k\in Z} h_{0k} w_n(2t-k) \\ w_{2n+1}(t) = \sqrt{2}\sum_{k\in Z} h_{1k} w_n(2t-k) \end{cases} \qquad (2)$$

in which h_{0k} is the db2 low filter and h_{1k} is the

db2 high filter because there's some fast wavelet algorithm available [9],then the approximation coefficients will be saved,while the detail coefficients is discarded. In this way, it can decrease the time of processing and the storage space of PCA training data significantly under the premise that the recognition rate isn't reduced, when dealing with huge amount of training faces.

After pre-processing the face image, PCA will be used to extract the orthogonal basis vectors and the corresponding eigenvalue from the set of training face images. PCA is one of the most popular method to face recognition by projecting data form a high-dimensional space to a low-dimensional space. In 1987 Sirovich and Kirby used Principal Component Analysis (PCA) in order to obtain a reduced representation of face images [13]. In essence, PCA is an optimal compression scheme that minimize the mean square error between the original images and their reconstructions[14,15]. To perform PCA, each of the training images should be the same size. Let's denote $A=[T_1...T_M]$ as the training set of faces in which each column represents a face image and then substract the mean face $\frac{1}{M}\sum_{n=1}^{M}T_n$ from each column

$$X_i = T_i - \frac{1}{M}\sum_{n=1}^{M}T_n \tag{3}$$

After that, extract the eigenvector(eigenface) and the eigenvalue from the covariance matrix $C=XX^T$. Because of the computationally intensive in covariance matrix, singular value decomposition is used in $X=UEV^T$, where U contains the eigenvectors of the covariance matrix C and then each training image will be projected into the eigenspace as a weight $Y_i=U^TX_i, i=1,...,M$ .In this way, the feature value $[Y_1\cdots Y_M]$ of the training faces can be obtained.

To find the closest match and improve the recognition rate further, SVM is used as the classifier instead of neural network [11]. In order to establish the optimal SVM classifier, each instance of the face feature in the training set should contains one class label and several feature values. When being tested, SVM model will predict the class label of the test-face which are given only the feature values. Suppose we are given a training set of instance-label pairs $(x_i,y_i), i=1,...,l$ where $x_i \in R^n$ and $y \in \{1,-1\}^l$, to find a linear separating hyperplane with the maximal margin in this higher dimensional space, the following optimization problem

$$\min_{w,b,\xi}\left(\frac{1}{2}w^Tw + C\sum_{i=1}^{l}\xi_i\right) \tag{4}$$

should be resolved which is subject to

$$y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \xi_i \geq 0 \tag{5}$$

In our method, we take the radial basis function(RBF):

$$K(x_i,x_j)=\exp(-r\parallel x_i-x_j\parallel^2), r>0 \tag{6}$$

as the kernel function, since unlike the linear kernel it can handel the case when the relation between class labels and feature values is nonlinear, then use cross-validation to find the best parameter C and γ for preventing the overfitting problem. After pre-processing, feature extraction and classification, the target face will be tracked by the algorithm supported by mean-shfit and kalman filter.

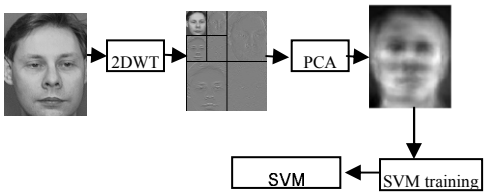

Figure 4. face matching framework

IV. TRACKING FACE WITH MEAN-SHIFT AND KALMAN FILTERING

When in the complex background, mean shift will fail to track the target face caused by partly covered or fast movement, as meanshift tracks the face only by the face color without movement prediction [10], it's very likely to lose track of the target face. On the purpose of improving robustness, the two dimensional Kalman filter is utilized to predict the movement and position of the moving target.

As an appearance based tracking method, the

meanshift tracking algorithm updates the weight of each pixels in the region and employs the meanshift iterations to find the candidate target which is the most similar to a given model in terms of intensity distribution[3], with the similarity of the two distributions being expressed by a metric based on the Bhattacharyya coefficient [1]. Given the position of the center of target face is y, and the candidate position might be $\{x_1, x_2, ..., x_n\}$, So the target candidate density could be

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_k} k(\| \frac{y - x_i}{h} \|^2) \quad (7)$$

and the target density is

$$\hat{q}_u = C \sum_{i=1}^{n} k(\| x_i^* \|^2) \quad (8)$$

where the $C = \frac{1}{\sum_{i=1}^{n} k(\| x_i^* \|^2)}$ and h is the bandwith of kernel function. With $\hat{q}_u$ and $\hat{p}_u$, Bhattacharyya coefficient can be represented as

$$\bar{\rho}(y) = \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(y)\, \hat{q}_u} \quad (9)$$

and the similarity distance can be $d(y) = \sqrt{1 - \rho[p(y), \hat{q}]}$. By translating the window in the direction of meanshift vector, the target could be tracked by the algorithm, but it still have some limitations in the tracking.

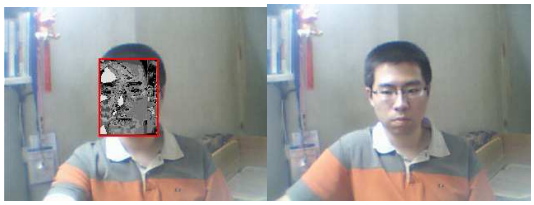

Figure 5. the weight of each pixels in the meanshift tracking region

In order to improve the robustness, two dimentional Kalman filter can be used to predict the starting position for mean shift iteration in the k+1 frame on the basis of the center position of the target in the k frame. Let's denote $A = \begin{matrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{matrix}$ (Δt is interval between frames) as the state transition matrix and $w(k)$ as system perturbation, $v(k)$ be the noise, $H = \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{matrix}$ be the observability matrix.

Therefore, the state of the system can be modeled as

$$x(k) = A * x(k-1) + w(k) \quad (10)$$

and the corresponding observation equation is

$$y(k) = H * x(k) + v(k), \quad (11)$$

in which $x(k) = [x, y, dx, dy]^T$ is the state vector at time k and $y(k) = [x, y]^T$ is the observed value, So the state vector $x(k+1)$ can be updated using the system model and measurement model.

When the object moves around, the size of the tracking window should be adaptive on the purpose of improving the tracking stability. LOG filter can be used to deal with this problem. As mentioned in[16], we substitute the LOG filter with DOG filter to reduce the processing time.

$$DOG(x;\sigma) = G(x;\sigma) - G(x;1.6\alpha) \quad (12)$$

Figure 6 left::remote range face

Right::near range face

The above images are the result of DOG filter. When the face approaches the camera, the sum of the gray scale in the tracking blob will increase and vice versa. Thereby we can change the sizeof tracking blob based on the change rate of the gral scale of the DOG result. In this way the tracking window can adjust its size, when the face approaches or moves away.

V. EXPERIMENTAL RESULTS

All the experiments are executed on a computer with 1.73Ghz pentium M processor and 512Mb ram in the windows xp system; besides, all the algorithm program is implemented in c++ code with VC 6.0.

Face detection experiment

We use MIT CBCL face databases as the face detecting training set which consists of 2429 faces

and 4548 non-faces 19 by 19 pixels grayscale images.

Figure 7 training face and two features of weak classifier.

As skin detecting is added to the face detection, it will take more time to detect the faces compared with original adaboost algorithm. On the purpose of reducing the time for detecting skin, the skin image is converted to an integral image. By doing this, the detecting time for processing a 320 by 240 pixel image can be reduced from 325ms to 87ms. Although detecting time is little bit longer than that of adaboost detector, there's a great improvement in the correct detection rate.

Figure 8. output of adaboost

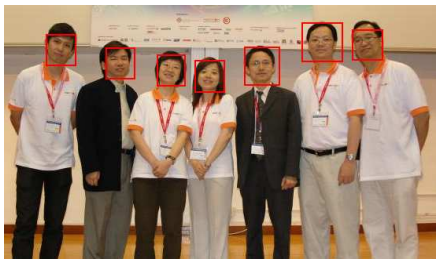

Figure 9. output of out detector

Face recognition experiment

we use the ORL face databases, within which there are 400 face images(112×92) of 40 people with different gesture and expression under different illumination condition.

Figure 10. ORL face images

We pre-process the image with the wavelet filter db2 and then extract 65 feature values in each image. In the SVM, as discussed in section 3,RBF is adopted as the kernel function.

$$K(x_i, x_j) = \exp(-\frac{\| x_i - x_j \|^2}{2\sigma^2}) \quad (13)$$

while $\sigma = 0.5$, C = 0.03125. Four face images in each face image set are selected to compose the face training set and the left six images of each person are utilized as the testing set. The result of experiment is listed in table 1 and figure 10.The time required for recognizing each face cost about 85ms.

TABLE 1: THE ACCURATE RATE OF TWO ALGORITHM UNDER DIFFERENT NUMBER OF SAMPLES IN EACH CLASS.

The number of samples in each class	5	4	3	2
Wavelet+PCA+SVM	96.0%	96.0%	92.0%	87.0%
Fisher	92.0%	91.0%	87.0%	82%
PCA	88.0%	87.0%	85.0%	78.0%

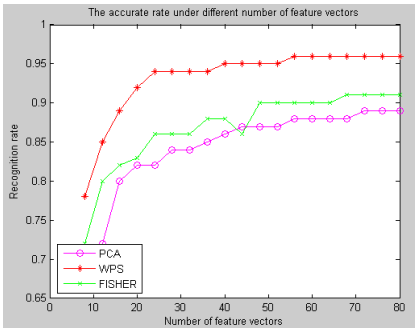

Figure 11. the accurate rate under different number of feature vectors

Face Tracking experiment

As for tracking face, the following pictures indicate the tracking performance in realtime when moving fast and partly covered. The time for tracking target face in each frame is about 32ms with 34 percent of full cpu load.

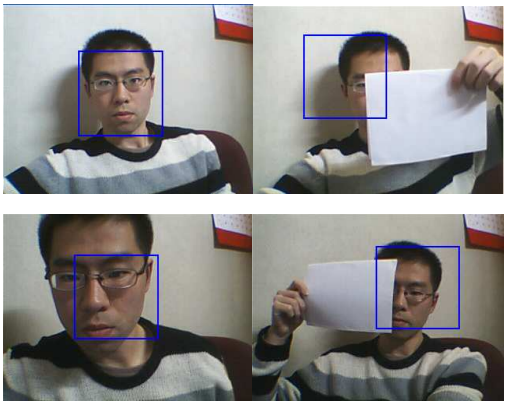

Figure 12. Tracking face partly covered

Figure 13. Tracking moving face by only meanshift

Figure 14. Tracking moving face by meanshift and kalman filter

Compared with the conventional meanshift algorithm, we can conclude that our algorithm is much more robust especially when the face is moving fast and partly covered.

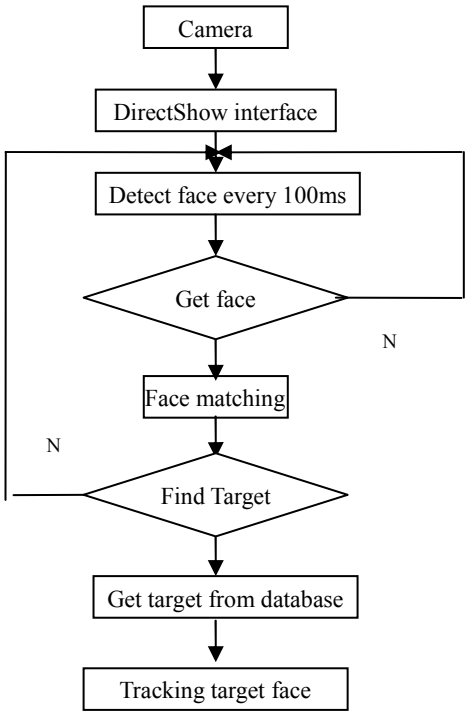

Figure 15. Program flowchart

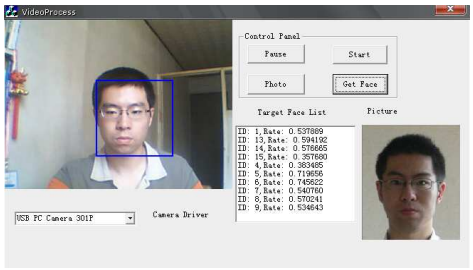

Figure 16. the output of the our program

VI. CONCLUSIONS

In this paper, we propose a realtime face recognizing and tracking system which can detect face, then recognize and track it. Our method performs well regardless of whether the faces is in the complex background, moving fast or partly covered, and the hybrid algorithm Wavelet/PCA/SVM further enhance the accuracy of face recognition. By experimental results, we have demonstrated that the proposed method dramatically improves the robustness and accuracy of the real-time face recognizing and tracking system

REFERENCES

[1] D. Comaniciu, V. Ramesh. 2000. Mean Shift and Optimal Prediction for Efficient Object Tracking, IEEE Int'l Conf. Image Processing, Vancouver, Canada, Vol. 3: 70-73.

[2] D. Comaniciu, V. Ramesh. 2000.Robust Detection and Tracking of Human Faces with an Active Camera, IEEE Int'l Workshop on Visual Surveillance, Dublin, Ireland, 11-18.

[3] D. Comaniciu, V. Ramesh, P. Meer. 2000. Real-Time Tracking of Non-Rigid Objects using Mean Shift, To appear, IEEE Conf. on Comp. Vis. and Pat. Rec., Hilton Head Island, South Carolina.

[4] Chien J T, Wu C C. 2002. Discriminant waveletfaces and nearest feature classifiers for face recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(12): 1644-1649.

[5] Yang Minghsuan, Kriegman D J, Ahuja N. 2002. Detecting faces in images-a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(1): 34-58.

[6] Viola P, Jones M. 2004. Robust Real-Time Face Detection[J] International Journal of ComputerVision 57(2): 137-154.

[7] Douglas Ngan King N. Face. 1999. Segmentation Using Skin-color Map in Videophone Applications[J]. IEEE Trans on CS, 9(4): 551-564.

[8] C.W.Hsu and C.J.Lin. 2002. A simple decomposition method for support vector machines.machine Learning, 46: 291-314.

[9] Huluta, E.; Petriu, E.M.; Das, S.R.; Al-Dhaher, A.H.;Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE Volume 2, 21-23 May 2002 Page(s): 1537 - 1542 vol.2

[10] S.J.McKenna, Y.Raja, S.Gong. 1999. Tracking Colour Objects using Adaptive Mixture Model, Image and Vision Computing,17: 223-229.

[11] G. Dai and C. L. Zhou, "Face Recognition Using Support Vector Machines with the Robust Feature", Proc. of the 2003 IEEE International Workshop on Robot and Human interactive Communication, 2003.

[12] Viola P, Michael J. Rapid Object Detection Using a Boosted Cascade of Simple Features[C]. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA. 2001.

[13] Sirovich L., and Kirby M. 1987 . A low-dimensional procedure for the characterization of human faces, J.Opt. Soc. Amer. A, vol.4, no. 3, pp. 519-524..

[14] P. Belhumeur, J. Hespanha, D. Kriegman. 1997. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence , 19(7):711–720

[15] M. Kirby, L. Sirovich. 1990. Application of the Karhunen-Loeve Procedure and the Characterization of Human Faces. IEEE Transactions on Pattern Analysis and Machine Intelligence , 12(1):103–108.

[16] R.T.Collins. 2003. Meanshift Blob Tracking through Scale Space. IEEE Conference on Computer Vision and Pattern Recognition, Vol.2, pp. 234-240.

Received: June 2009. Accepted: Sep. 2009