Face Segmentation Based on Skin Color in Complicated Background and Its Sex Recognition

Chuanxu Wang
Faculty of Informatics Qingdao University of Science & Technology, Qingdao, China
E-mail: wangchuanxu_qd@163.com

***Abstract*—Human skin color distribution is relatively compact in a color space, and it changes from frame to frame following illumination variations, also it is subject to camouflage interruption in complicated background for skin segmentation. This paper mainly talks about face detection based on skin color feature in complicated background. We assume that skin pixels in each frame are closed together as a "dot cloud" in a color space, its shape evolution from frame to frame is modeled as the mixture of translation, scaling and rotation. We introduce linear combination of forecasts to predict these parameters related to the above shape evolutions, so that skin distribution of next frame to be segmented can be predicted and skin segmentation for face detection can be improved against illumination variations. Additionally human skin biological feature is then introduced to remove camouflage noise. For face gender recognition we adopt 2-D Gabor transform for feature extraction and SVM for recognition. Extensive tests prove that this algorithm is quite sensitive to human color, and more accurate for human skin segmentation with Bayes classifier; Good performance of gender classification test is also achieved on a relative large scale and low-resolution video database.**

***Index Terms*—skin segmentation, combination of forecasts, gender classification, 2-D Gabor-wavelet**

I. INTRODUCTION

In video surveillance system, big-scaled information of human body movement is preferably useful for supervision in far distance situation, such as gesture recognition and gait recognition; while for supervision in near distance situation we need high resolution information which can often be used for face recognition or hand movement recognition [1].

A. Skin Color Detection

Human skin color is an eminent feature for face segmentation, and independent on texture and also different from most of background [2]. Many skin segmentation methods are become invalid when illumination changes and camouflage disturbs. But only a few methods adopt color stabilization and dynamic adaptation. [3-5] used color stabilization technique and threshold value segmentation or ellipse classifier for human skin detection, which could adapt illumination variation to some extent, but they only used two chromatic vectors and ignored brightness information, which incurred low accuracy of segmentation. [6-9] adopted skin area tracking technique which lost its color tracking caused by background noise and finally these methods wrongly trapped at none skin area. The crux of skin area tracking in video frames is to predict its movement before tracking and segmentation. Therefore an improved human skin segmentation method is proposed, we assume skin patch evolution in the color space from frame to frame is parameterized as translation, scaling and rotation. The linear combination of forecasts, which is consisted of 2-order Markov predictor and Wiener one step predictor, is proposed to predict these evolution parameters. These predictions will foretell the skin distribution in next frame, which help to get high accuracy of skin segmentation.

B. Gender Recognition

Face image gender recognition is an extensively concerned problem in computer vision. Early gender recognition in computer vision was mainly based on machine learning of neural network. Gollomb [10] introduced a two-layered neural network SEXNET, its input face image was 30×30, and the classification accuracy was 91.9% tested with 90 face images that were 45 man face images and 45 woman face images respectively; Another neural network designed by Edelman [11] was trained with 3 different part of face images, which were top half part and bottom part of face image and the whole face image. Linear nerve cells were used inside, and 3 different classification results were compared. Tamura's neural network [12] was consisted of multiple layers, and gender recognition test was carried out on face images of different resolutions, which got 7% error recognition rate for 8×8 ultra low resolution face images. In short, all these neural network methods need large amount of training samples but result in low generalization; further more its recognition speed is quite slow. In the recent years Baback Moghaddam and M.H.Yang [13] begin to adopt SVM (Support Vector Machine) for gender recognition, and have made recognition comparisons among RBF network and FLD and Minimum Vicinity classifiers based on FERET face image database, which conclude that SVM classifier performs the best. But these tests are conducted with gray image directly which are not robust to illumination variation and of high dimensioned data. Therefore PCA is used in some methods [9,14] to decrease its dimension, but PCA ignores information among different patterns [15]. In order to overcome the above shortcomings, 2-D Gabor wavelet transform is used instead to extract face image feature in this paper and SVM is our choice for recognition.

doi:10.4304/jsw.6.7.1209-1216

C. Orgnizition of This Paper

The rests of the paper are organized as following. Section II gives an overall description of the segmentation methods based on Bayes classifier. In section III we give the model of skin color distribution variation from frame to frame while the illumination changes, and introduce the combination of forecasts algorithm to predict this variation. We detail face feature extraction with 2-D Gabor transformation and SVM training for gender recognition in section IV. Experiments are conducted for both face patch segmentation and gender recognition in section V to test the effectiveness and robustness of our method. Finally we give conclusions and future work in section VI.

II. SKIN SEGMENTATION BASED ON BAYES CLASSIFIER

Suppose a pixel color is rgb, its conditional probabilities of $P(rgb|fg)$ and $P(rgb|bg)$ show its occurrence possibility in foreground (skin area) and background (none skin area); while conditional probabilities of $P(\text{fg}|rgb)$ and $P(\text{bg}|rgb)$ present it belonging to skin (foreground) and none skin (background) respectively. Their ratio is called classification threshold.

That is

$$K < \frac{P(fg|rgb)}{P(bg|rgb)} = \frac{P(rgb|fg)P(fg)}{P(rgb|bg)P(bg)} \tag{1}$$

Then

$$K \times \frac{1-P(fg)}{P(fg)} < \frac{P(rgb|fg)}{P(rgb|bg)} \tag{2}$$

Here $P(fg)$ is the probability of an arbitrary pixel in an image being skin. K can be determined empirically on the consideration, which the classification accuracy is at least 85% and fault alarm rate is less than 25%, K finally is decided as 0.0673 [16] . So each pixel can be classified as pixel or not via formula (2), which is called Bayes classifier.

III. VARIATION PREDICTION OF SKIN DISTRBUTION CAUSED BY ILLUMINATION CHANGE

A. Evoution Modeling for Skin Pixel Distribution

Study of [15, 17] shows that skin color distribution is compact in a color space, we call this skin pixel collection as "dot cloud", and its distribution changes frame by frame as illumination varies, also this evolution per frame is smooth. In order to simply model this evolution, some assumptions are hypothesized, that is, there is only global change and no local deformation in the evolution, this progress can be parameterized as translation and rotation and scaling, which can be described with 8 variables and shown as Fig 1. The parameters of translation and scaling can be respectively represented with 3 means and 3 variances of each color vectors of skin pixels in "dot cloud", while rotation can be described with 2 angle variables as θ and ϕ in Fig 2.

Vectors of $e_{1,\tau}$ and $e_{2,\tau}$ and also $e_{3,\tau}$ are eigen vectors of

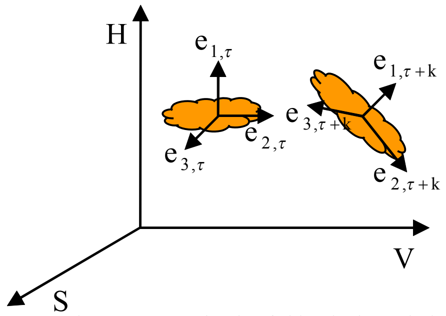


Figure 1. "Dot cloud" of skin pixels evolution from frame to frame.

covariance matrix of skin pixels in frame τ , while $e_{1,\tau+k}$ and $e_{2,\tau+k}$ as well as $e_{3,\tau+k}$ are the same for

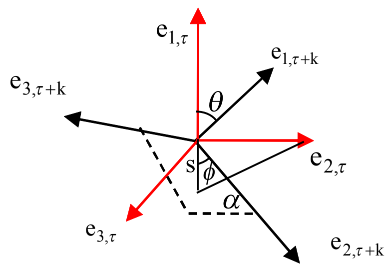


Figure 2. 2 Definition of Angle Parameters of "Dot cloud" evolution from frame to frame.

frame $\tau+k$. θ is the angle between $e_{1,\tau}$ and $e_{1,\tau+k}$, and ϕ is the angle between $e_{2,\tau+k}$ and s , which s is the projection of $e_{2,\tau}$ towards the plane formed by $e_{2,\tau+k}$ and $e_{3,\tau+k}$.

B. Linear Combination of Forecasts for Parameter Prediction of "Dot Cloud" Evolution

The shape of "dot cloud" implies the distribution of skin pixels in a frame. For a new frame to be segmented if its shape of "dot cloud" can be predicted, then its skin pixel histogram and none skin pixel histogram can be figured out, thereafter the probability $P(\text{fg}|rgb)$ and $P(\text{bg}|rgb)$ can be updated, which can improve the skin segmentation accuracy according to Bayes classifier, when illumination changes. So the key problem is to forecast this shape evolution from its past shape parameters. In this paper linear combination of 2-order Markov predictor and Wiener one step predictor is proposed as following.

C. Two Algorithms for Parameter Prediction

[18-19] show that the above mentioned "dot cloud" evolution from fame to frame is a stationary stochastic process, so the parameters of shape evolution can be predicted with Wiener one step predictor.

Suppose the past data sequences are $x(n-1)$, $x(n-2)$, …, $x(n-p)$, and then $x(n)$ can be predicted as

$$\hat{x}(n) = -\sum_{k=1}^{p} a_{pk} x(n-k).$$

Where a_{pk}, $k = 1, 2, \cdots, p$ can be figured out by Yule-walk equation:

$$\begin{bmatrix} \phi_{xx}(0) & \phi_{xx}(1) & \cdots & \phi_{xx}(p) \\ \phi_{xx}(1) & \phi_{xx}(0) & \cdots & \phi_{xx}(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{xx}(p) & \phi_{xx}(p-1) & \cdots & \phi_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{p1} \\ \vdots \\ a_{pp} \end{bmatrix} = \begin{bmatrix} E[e^2(n)]_{\min} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3).$$

$E[e^2]$ is mean square error defined $E[e^2] = E[x(n) - \hat{x}(n)]^2$.

Another frequently used predictor is 2-order Markov predictor model [20-21]. it is described as

$$x_{n+1} = A_0 x_{n-1} + A_1 x_n + B\omega_n.$$

where parameters A_0 and A_1 stands for deterministic parts, which show the correlation of consecutive frames; while coefficient B are stochastic part, which can not be predicted and controlled. For the past data of m frames, A_0 and A_1 can be figured out by MLE (Maximum Likelihood Estimation) theory with $m-2$ historical data, they are the solution of the following equation:

$$\begin{cases} S_{20} - \hat{A}_0 S_{00} - \hat{A}_1 S_{10} = 0 \\ S_{21} - \hat{A}_0 S_{01} - \hat{A}_1 S_{11} = 0 \end{cases} \quad (4)$$

Where

$$S_{ij} = \sum_{n=1}^{m-2} X_{(n-1)+i} X_{(n-1)+j}^T \qquad i,j = 0,1,2$$

And

$$\hat{C} = \frac{1}{m-2} Z(\hat{A}_0, \hat{A}_1)$$

Where

$$Z(A_0, A_1) = S_{22} + A_1 S_{11} A_1^T + A_0 S_{00} A_0^T - S_{21} A_1^T - S_{20} A_0^T + A_1 S_{10} A_0^T - A_1 S_{12} - A_0 S_{02} + A_0 S_{01} A_1$$

While $C = BB^T$, then B can be made out.

D. Linear Combination of Forecasts

None of any prediction algorithm is always perfect, but their combination can achieve a better estimation statistically. Suppose $f_{t,i}$ $i=1,2$ stands for the i^{th} predictor model at time t, then the linear combination of these two models is $f_t = k_1 f_{t,1} + k_2 f_{t,2}$, where k_i represents the weight of i^{th} model; also suppose $e_t = \sum_{i=1}^{2} k_i e_{t,i}$ as its prediction error, where $e_{t,i} = f_{t,i} - y_t$ $(i=1, 2)$ is absolute prediction error of the i^{th} prediction model at time t. Therefore prediction accuracy information of this combination model is kept in error sequence e_1 and e_2. The MSSE (Minimum Sum of Square Error) principal is adopted to determine $f_{t,i}$ $i=1,2$ with the past N data. That is

$$\mathrm{SSE} = \frac{1}{N}\sum_{t=1}^{N} e_t^2 = \frac{1}{N}\sum_{t=1}^{N}\left[\sum_{i=1}^{2} k_i e_{t,i}\right]^2 = \frac{1}{N}\sum_{t=1}^{N}\left[\begin{bmatrix} k_1 & k_2 \end{bmatrix}\begin{bmatrix} e_{t,1} \\ e_{t,2} \end{bmatrix}\begin{bmatrix} e_{t,1} & e_{t,2} \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}\right]$$

$$= \frac{1}{N}\sum_{t=1}^{N}\left[\begin{bmatrix} k_1 & k_2 \end{bmatrix}\begin{bmatrix} e_{t,1}^2 & e_{t,1}e_{t,2} \\ e_{t,1}e_{t,2} & e_{t,2}^2 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}\right] = \frac{1}{N}\sum_{t=1}^{N}\left[\begin{bmatrix} k_1 & k_2 \end{bmatrix} A_m \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}\right] = \frac{1}{N} K^T A_m K$$

Shorten as:

$$\min(SSE) = \frac{1}{N}\min(K^T A_m K)$$

$$s.t. \begin{cases} \sum_{i=1}^{2} k_i = 1 \\ k_i \geq 0 \quad i = 1,2 \end{cases}$$

Its solution is

$$K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}.$$

IV. FACE FEATRUE EXTRACTION AND SVM TRAINING FOR GENDER RECOGNITION

Video face segmentation based on skin color can be achieved with our method in complicated background. But it is not suitable as input signal for gender recognition directly, because it is not robust to image rotation and partial occlusion and illumination as well. Therefore we adopt 2-D Gabor transformation for feature extraction, and then we train SVM for gender recognition.

A. Face Feature Extraction With 2-D Gabor

Tai Singlee [22] extended 1-D compact wavelet to 2-D Gabor wavelet in 1996. Its application on face image recognition is originated that coefficients of 2-D Gabor transform are of better visual characteristics and biological background. Its advantages are as following, better time-frequency localization and better signal resolution in time and frequency domain, and especially it is robust to illumination variation and adaptable to image deformation and rotation to some extent. In all, Gabor transform is better than others for gender biological characteristics extraction.

2-D Gabor core function can be defined:

$$\varphi_{u,v}(x, y) = \frac{k^2}{\sigma^2}\exp\left(-\frac{k^2(x^2+y^2)}{2\sigma^2}\right)\cdot\left[\exp(ik\cdot(x,y)^T) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \quad (5)$$

where i is a complex operator; (x, y) is a spatial position; parameter σ / k decides the size of Gaussian window, and let σ equal $\sqrt{2}\pi$; k is the wave vector, defined as

$$k = \begin{pmatrix} k_x \\ k_y \end{pmatrix} = \begin{pmatrix} k_v \cos\theta_u \\ k_v \sin\theta_u \end{pmatrix}.$$

And $k_v = 2^{-\frac{v+2}{2}}\pi$, $\theta_u = u \cdot \frac{\pi}{K}$; frequency coefficient v determines Gabor filter wavelength; while direction coefficient u decides the orientation of Gabor core function, k_v defines the central spatial frequency of

Gabor filter. The second item in square brackets of formula (5) is direct current which is subtracted, so that Gabor is robust to illumination variation; $\exp(-\frac{k^2(x^2+y^2)}{2\sigma^2})$ is Gaussian function, which limits the bound of oscillation, and makes it valid in a local area, so Gabor filter is adaptable to image deformation somewhat.

In this paper we discrete $v = 0,1,2,3$ and θ $\theta = 0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}$. So there are 32 Gabor wavelets for human face feature extraction.

Gabor wavelet transform is realized via convolution computation between face image and each of 32 different Gabor core functions respectively, which is:

$$G = \varphi_{u,v} * I(x,y) = \iint \varphi_{u,v} I(x,y)dxdy \tag{6}$$

Herein, $I(x,y)$ is gray face image, Convolution result is complex, so its absolute value can be regarded as Gabor transform result.

B. SVM Training For Gender Recognition

Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of support vector machines (SVM), a data point is viewed as a p-dimensional vector, and we want to know whether we can separate such points with a $p-1$-dimensional hyperplane. This is called a linear classifier. There are many hyper planes that might classify the data. One reasonable choice as the best hyper plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper plane exists, it is known as the maximum-margin hyperplane and the linear classifier it defines is known as a maximum margin classifier. For none linear classified data set, a transform can be carried with a nonlinear kernel function. This allows the algorithm to fit the maximum-margin hyper plane in a transformed feature space. The transformation may be nonlinear and the transformed space high dimensional; thus though the classifier is a hyper plane in the high-dimensional feature space, it may be nonlinear in the original input space.

We use BioID face database as training samples. There are 976 men and 544 women face images, some are from different illumination, different gesture, different facial expression and different face making up (glass on/off, beard or non-beard etc.), image size is 286×384.

Face images should be normalized before feature extraction. They are manually labeled out the positions of eyes and mouths and enhanced with histogram equilibration method and rotated and also cut off background parts, finally face images are normalized as 21×18 pixels. One example of normalized face image is shown in Fig 3.

Gabor transform is carried out on these normalized images and face features are extracted to form training vectors for SVM. There are 32 Gabor wavelet filters with different frequencies and directions, so one input face will have 32 times of convolutions and 32 output images. These 32 real parts and norm images are shown in Fig.4.

(a) (b) (c)

Figure 3. Face image normalization
(a) Original image; (b) manually located positions of eyes and mouth for normalization (c) Normalized image.

So SVM training data are increased to a great deal, in order to reduce SVM training time, we propose to calculate the variances of each column in a Gabor norm image and the mean value of the whole image, so 32 feature images of an input face image will become a smaller feature vector of 1×608. In this paper, 2620 such vectors are used for SVM training.

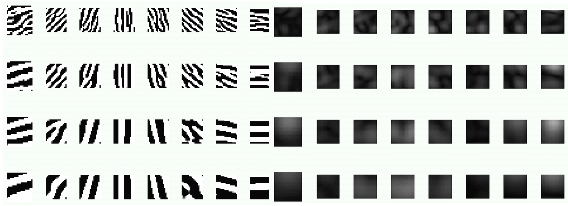


Figure 4. Real part and norm images of 32 Gabor outputs

Finally face image gender recognition tests are performed with SVM on the rest samples of this database.

V. Tests and Analysis

There are three parts of test, the first experiment is validity verification of linear combination of forecasts, which we will compare Wiener predictor and Markov 2-order predictor and their combinations. The second part is about face segmentation, in which we will show the illumination variation influence on skin segmentation and test the robustness of our method. The final test is gender recognition on face image after skin segmentation. All these three experiments are conducted with Matlab 7.01 and CPU 2.33 GHz with 4G Megabytes memory. They are illustrated as following respectively.

A. Test of Linear Combination of Forecasts

Data prediction is on the basis of its intrinsic self-correlation. So first of all we should design an auto correlated signal as our test sample. Any signal of rational power spectrum can be regarded as the output of a linear system, whose input is a white noise signal. We produce a white noise signal as input to an AR (Auto-Recursive) filter, and then its output is an auto-correlated stochastic sequence.

Three prediction models are test respectively, which are Wiener one step model and 2-order Markov model and their linear combination model. Their prediction results are compared and illustrated in Fig.5 and their error variance and means are compared in Tab.1.

In Fig.5 there are two group of prediction result charts illustrated as group A and B. The solid line in Fig.5 represents test sample while the dashed line stands for predicted result. Through the simulation tests, the conclusion can be drawn that linear combination of

(A)

(B)

Figure 5. Prediction comparisons in group A and B. In each group from top to bottom are Wiener prediction and Markov prediction and linear combination of predictors respectively.

TABLE I.
COMPARISON OF ERROR VARIANCE AND MEANS OF THE 3 PREDICTORS

Test group \ Predictors		Wiener prediction	Markov prediction	Linear combination
Test A	Error variance	0.985652	0.223291	0.177097
	Error means	0.042086	-.013956	-0.003145
Test B	Error variance	1.116644	0.187231	0.131621
	Error means	-.009779	0.002905	0.000470

predictions is better than any single model statistically; in that prediction error variance of linear combination model is less than Markov model and Wiener model.

B. Skin Segmentation Test Against Illumination Variation and Background Camouflage Noise

Skin segmentation tests are conducted in two aspects, which are robustness test agianst illumination vairation and skin camouflage interruption. They are analyzed as following.

1) *Skin segmentation test against illumination variation*
Illumination variation can cause skin color distribution modified frame by frame. In order to improve the skin segmentation accuracy with Bayes classifier it is necessary to predict skin distribution variation.

In Fig 6, picture **a** is the original image, where there are many lights shining and illumination varies from time to time, therefore human skin distribution changes frame by frame. So this variation should be predicted and updated before Bayes segmentation, otherewise skin detection will be mistaken. Linear combinations of forecasts are

Figure 6. Skin segmentation test against illumination variation. From top to bottom pictures are named as **(a) (b) (c) (d) (e)**.

used, its result is shown with picture **c,** while picture **b** is the result of none prediction method. Comparison between picture **b** and **c** shows that the linear combinition of forecasts is better than none prediction method, which has less misclassification pixels. Picture **d** is the morphological process result of **c**, and **e** is the corresponding skin patch from original image.

2) *Skin segmentation test against camouflage*
Background camouflage interference is one of the most difficult problems for human skin detection. In this paper human skin biological feature is adopted to remove

Figure 7. Background camouflage noise removing tests, from top-lest to bottom-right pictures are named as **(a) (b) (c) (d) (e) (f).**

camouflage noise. That is, for human skin pixel, its color is (r,g,b), there are more red components than blue and green parts, their ratio is $1.1 \le r/g \le 1.5$, $1.3 \le r/g \le 1.7$, $1.0 \le g/b \le 1.3$.

In our test video there is a yellow package which has similar color to human skin in Fig.7, picture ***a*** is an original frame. Comparisons of picture ***b*** and ***d*** are segmentation results produced by linear combination of forecasts and none prediction method respectively. Background color camouflage "yellow package" is partially misclassified in both pictures. It is obvious that picture ***b*** is better than ***d***, which has less misclassification pixels. Map ***c*** is corresponding part of ***b*** in frame ***a***, Human skin biological feature is also used to filter out camouflage noise such as ***e***, and picture ***f*** is face image which is corresponding to the mask picture ***e*** in picture ***a***.

C. Gender Recognition Test and Analysis

There are two schemes for gender recognition test. In order to test the validity of Gabor transform method for feature extraction, our test is carried out with different SVM core functions for both pixel-based image which is not transformed with Gabor wavelet and the transformed image; another scheme is to use different sized face images to make out how face image size influence recognition accuracy. They are illustrated as follows.

1) *Gabor feature extraction for gender recognition*

Validity test for face feature extraction in gender recognition with Gabor transform is accomplished with 120 face images. The results are shown with Tab.2, in which the third column represents the wrongly classified face images among 120 test samples, and number pair 27/28 stands for Gabor method versus pixel-based image method, and it is true for the rest number pairs in column 3 and 4.Through the data comparisons, gender recognition accuracy for Gabor transformed images is obviously better than that of pixel-based images when using polynomial and linear core functions of SVM; and for the core functions of RBF and Sigmoid core functions the correct recognition rate is the same for both Gabor transformed images and pixel-based images. Conclusively the Gabor method for face feature extraction in gender classification is valid and outperforms than pixel-based images and the best parameter fitting is polynomial core function with parameter d=3.

TABLE.II
VARALITY TEST OF GABOR TRASFORMATION WITH DIFFERENT CORE FUNCTIONS OF SVM（C=32）

Core function	Parameters for core function	Error classified images	Correct ratio
Linear core	\	27/28	77.5%/76.67%
Polynomial core	d=3	15/20	87.5%/83.33%
RBF core	Gamma=0.0078125	30/30	71.67%/71.67%
Sigmoid core	Gamma=0.0078125	30/30	71.67%/71.67%

2) *Influences of different normalized face sizes on gender recognition*

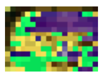
In this experiment our purpose is to verify the how the face size influence the gender recognition accuracy. Different sized face images of 21×18 and 35×30 and 49×42 are as tested and their results are compared. First of all face images are segmented from different videos and then normalized to different sizes for testing. Our face segmentations are from three different videos, they are detailed as followings respectively.

The first face segmentation test is from a video with simple background but there are two faces in each frame. Its segmentation result is shown in Fig.8. Each face image is normalized as 21×18.

Figure 8. Segmentations for two face images in each frame from a simple background.

The second test video is with complicated background but only one face in each frame in Fig.9. There are camouflage noise to skin color, such as yellowish closet and yellow cover of an electrical outlet in the wall. We use our skin color predictor to improve the skin segmentation and adopt skin biological feature to remove camouflage noise and then morphological filters are carried out to remove small sized noise patches and fill the small holes in face image. This process is shown in pictures (A) of Fig.9; these faces are normalized as size 35×30, which are shown in pictures (B).

(A)



(B)

Figure 9. Face segmentation under complicated background. Pictures in (A) show the processing steps; pictures in (B) are normalized faces from frame 25 to 115 with 10 frame intervals, which are for Gabor transformation and gender test.

The third test video is similar as the second, but there are two faces in each frame, the processing steps are also similar to those of second video, they are illustrated in Fig.10, but face images are normalized as 49×42.

Figure 10. Segmentations for two face images in a frame from complicated background.

Different sized face images are firstly transformed with Gabor wavelet and then recognized with SVM. In this test polynomial core function is used and its parameter d is 3. The whole test frames are also 120, their test results are shown in Tab III.

TABLE III
CORRECT RECOGNITION RATE COMPARISON

Image Size	Number of error recognized images	Correct rate
21×18	15	87.5%
35×30	11	90.83%
49×42	9	92.5%

It is clear that correct recognition rate goes higher as face image size becomes large, but their training cost time also increases. So there is a good compromise between face image size and correct recognition rate. In this paper test shows that image size about 50 × 40 will offer 90% correct rate while the training speed is also acceptable.

VI. CONCLUSIONS AND FUTURE WORK

An improved method of human skin segmentation is introduced in this paper; its crux is parameter prediction. Through theory deduction and simulation tests it is proved that linear combination of forecasts method is better than any single prediction model, which improves skin detection sensitivity and has less misclassification pixels. Data prediction is on the basis of its self-correlation, which is a strong correlated stochastic process has better prediction accuracy than a less correlated one for the same predictor. This skin detection method can be used for human face detection, which can detect not only full-face but also side-face and occluded face.

Gender recognition for pixel-based image will be subject to illumination change, while big sized face image usually incurs high-dimensioned data and suffers from slow training speed. Gabor wavelet transform is proposed to extract face features and to keep robust to image gray variation, different facial expression and face ornaments. Finally SVM classifier is used for recognition for generalization improvement. The proposed algorithm is proved to be valid for face gender recognition in video, and of high recognition rate for frontal face image. Our future work is to speed up the whole algorithm for gender classification in real time, and also to improve skin detection accuracy for hand image segmentation and used for hand gesture recognition.

ACKNOWLEDGMENT

The author would like to express my thanks for sharing BioID Face DB. This research is partially supported by Natural Science Fund of Shandong (ZR2010FL007 and ZR2009GM007 and Y2008G09) and the Project of Shandong Province Higher Educational Science and Technology Program (J10LG23 and J09LG12).

REFERENCES

[1] P. Kakumanu, S. Makrogiannis, N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition,* vol.40, no.3, pp.1106 – 1122, 2007.
[2] Cheng Duansheng, Liu Kaisheng, "Summarization of skin detection techniques," *Journal of Computer*, vol.29 no.2 pp.194-207, Feb，2010.
[3] M. Soriano, J.B. MartinKauppi, S. Huovinen, "Adaptive skin color modeling using the skin locus for selecting training pixels," *Pattern Recognition* ， vol.36, no.3, pp.681-690, 2003.
[4] H.J. Anderson, E. Granum, "Tracking regions of human skin through illumination changes," *Pattern Recognition Letter*, vol.24, no.11, pp. 1715-1723, 2003.
[5] Li jie, Hao Xiaoli, "A Face Detection Method Based on Ellipse Model," *Computer Measurement and Control*, vol.14, no.2, pp.170-173, 2010.
[6] Zhu Q. , Cheng K. , Wu C. , Wu Y., "Adaptive learning of an accurate skin-color model," *In Proc. of 6th Conference on AFGR* , pp.37-42, 2004.
[7] Stern H. , Ef ros B. , "Adaptive color space switching for face tracking in multi-colored lighting environments," *In Proc. of 5th Conference on AFGR*, Washington, DC, USA, 2002, pp.249-255.
[8] C. Stauffer and W. Grimson, "Learning patterns of activity using real time tracking," *IEEE Trans. PAMI*, vol.22, no.8, pp.747-757, Aug. 2000.
[9] Wang Hui, "Application of PCA and SVM in human face recognition," *Computer Technology and Development*, vol.16, no.8, pp.24-27, Agu.2010.
[10] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," *Advances in Neural Information Processing Systems 3*, pp.572–577, 1991.
[11] B Edelman, D Valentin, H Abdi. "Sex classification of face areas ： How well can a linear neural network predict human performance", *Journal of Biological System*,1998,vol.6, no.3, pp.241-264.
[12] Tamura S ， Kawai H ， Mitsumoto H, "Male/female identification from 8×6 very low resolution face images by

neural network," *Pattern Recognition*, 1996, vol.29, no.2, pp.331-335.
[13] B Moghaddam, M H Yang, "Gender classification with support vector machines," *IEEE Trans. PAMI*， 2002. vol.24, no.5, pp.707-711.
[14] Xia Peng, Zhang Haoran, "Algorithm of increment PCA and its application for human face recognition," *Computer Engineering and Application*, 2010， vol.44, no.6, pp.228-231.
[15] Wu Bo and Ai Haizhou, "Human face gender classification," *Journal of Computer Research and Development,* vol.40, no.11, 2003, pp:1546-1553.
[16] M.J. Jones and J.M. Rehg, "Statistical color models with application to skin detection," *In Proc. CVPR,* vol.1, pp. 274-280, 1999.
[17] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Skin detection in video under changing illumination conditions," *In Proc. Int'l Conf. Pattern Recognition*, vol. 1, pp.839-842, 2000.
[18] J. Barron, D. Fleet, S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 42-77, 1994.
[19] A. Blake and M. Isard, *Active Contours*. Cambridge Univ. Press, 1998.
[20] Ma Yongkai, Tang Xiaowo, "Research of optimal problem on linear combination of forecasts," *System Engineering and Application*, 1998, vol.18, no.9, pp.110~114,123.
[21] Tang Xiaowo, *Prediction Ttheory and its Application*, Chengdu: Electronic University Press, 1992.
[22] Tai Singlee, "Image representation using 2D Gabor wavelets," *IEEE Trans. PAMI*, vol.18, no.10, 1996.

Chuanxu Wang received his B.Sc. in electronic engineering in 1990, M.Sc. in industry automation in 2000, both from Petroleum University, China, and PhD in industry automation in 2007 from The University of Ocean University, China. From Dec. 2008 to Oct. to Apr. 2009, he was a visiting researcher at AMRL at Wollongong University Australia. He is currently an associate professor in Informatics Institute, Qingdao University of Science and Technology, China. His research interests include signal processing and pattern recognition. He has published over 50 papers in refereed international conferences and journals.