Face Detection Based on Facial Features and Linear Support Vector Machines

Jinxin Ruan
College of Electronic and Information Engineering
South China University of Technology
Guangzhou, China
ruanjinxin@163.com

Junxun Yin
College of Electronic and Information Engineering
South China University of Technology
Guangzhou, China
eejyin@scut.edu.cn

***Abstract*—Face detection is a complicated and significant problem in pattern recognition and has wide application. This paper proposes a fast face detection algorithm based on facial features and linear Support Vector Machines (LSVM). First, using of skin color information, the algorithm quickly excludes most background regions from the images primarily leaving the skin color regions. Then we use LSVM to separate more non-face regions from the remaining regions, for exiting big differences between the face regions and non-face regions. Finally, we identify the face candidates by detecting eyes and mouth. The experimental results demonstrate that the algorithm can further improve the detection accuracy and lower false detection rate and greatly speed up the detection rate.**

Keywords-Face Detection; Skin Color Detection; Facial Feature; Support Vector Machines

I. INTRODUCTION

In many situations, it is crucial to quickly and accurately map a face in an image for face recognition, facial expression recognition, and other applications. In this paper, a rapid face detection algorithm based on facial features and LSVM is presented.

As an independent research field and with a promising future, face detection and recognition technology has attracted much study [1-2].

The skin color is an important feature of a face [3]. It has a strong cluster feature such as in the YCbCr and HIS color space. And, more importantly, it is not affected by body posture and facial expression giving it relative stability. Moreover, skin color is easily distinguished form most background color allowing background to almost be eliminated well in the preliminary detection by using skin color. At the same time, the scope of face detection is narrowed. The algorithm is simple, less computing, efficiency and easy to achieve real-time processing. Therefore, the face detection algorithm based on the skin color model is widely used [4]. But it is not sufficient to completely and accurately detect the face only by using the skin color information. It usually needs other facial features, such as information of the eyes, mouth or combines with other means. However, when we use the characteristics to verify the candidate regions, it is vulnerable to the skin color model used in the algorithm as well as the effect of skin color segmentation. Moreover, when several faces are very close to each other or the face region and other body regions or skin-likelihood background linked together, it often leads to false detection. This problem can be avoided by detecting the face candidate regions with statistical methods.

Because the number of face samples in face detection is very small, it is an issue of small samples. Statistical learning theory is currently considered to be the best theory for small samples statistics estimates and projections learning. SVM theory [5] is built on the basis of statistical learning theory; its purpose is to solve the problem of the learning and classification of small samples. With the SVM theory development, more and more attention is attracted and SVM is used in many applications [6]. [7] firstly applied SVM in face detection and achieved good results. However, non-linear SVM was used directly in the algorithm [7] to classify face region and non-face region. Due to the computing complexity of non-linear SVM, the algorithm is not for being used in practical application. Therefore, a Hierarchical Support Vector Machines [8] was introduced, most simple background regions are excluded from the image quickly by LSVM firstly, and then the rest candidate regions are confirmed by non-linear SVM. By this algorithm, the detection rate is improved with reducing false detections and higher detection speed. However, the algorithm does not make full use of the face information. In particularly, most of the non-face samples generate from gray scenery images. The non-face samples lack validity, lead to lower the performance and efficiency of the SVM classifier.

In this paper, we separate the skin color regions from the complex background in the YCbCr color space. And the LSVM classifier is designed to be trained with samples which are intentionally selected from the skin color regions. Then the most non-face regions can be rapidly ruled out from skin color regions. Compared with the traditional methods, samples are selected from the color images, and have rich useful information. Furthermore, non-face samples are extracted from the skin color regions. Therefore the effectiveness of samples is improved and easier for training. In order to avoid the computing complexity of non-linear SVM classifier, the information of eyes and mouth is used to verify the face candidate regions. Experiments show that this method can improve the detection accuracy and lower false detection rate, and greatly increase the speed of detection.

978-0-7695-3522-7/09 $25.00 © 2009 IEEE
DOI 10.1109/ICCSN.2009.76

II. SKIN COLOR MODEL

As the lighting condition, noise and so on has great effect on the skin color detection, lighting compensation and denoising must be carried out at first.

In the RGB color space, each component represents not only color but also luminance. The triple color components have great relevance each other and are vulnerable to lighting condition. So we adopt YCbCr color space which is the color space is popularly used in skin color detection. YCbCr color space is developed for TV systems, and it is a luminance separated color space which is widely used in MPEG, JPEG and other video compression standards. In the YCbCr color space, the chrominance of skin color has good color cluster feature, and it is simple to convert the RGB color space to the YCbCr color space. Moreover the color properties could be analyzed from compressed video stream directly without decompression. So the YCbCr color space is one of the most attractive color spaces in skin color modeling.

YCbCr color space could be got from RGB color space by linear conversion. But in order to further reduce the lighting effect and to obtain a better skin color cluster result, a segmented non-linear conversion [9] is used to convert the YCbCr color space into the $YC_b^{'}C_r^{'}$ color space. According to the elliptical cluster approach [9], we segment skin color regions refer to (1) in the $YC_b^{'}C_r^{'}$ color space, and attain a satisfactory result.

$$\frac{(x-ec_x)^2}{a^2}+\frac{(x-ec_y)^2}{b^2}<1 \tag{1}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b^{'}-c_x \\ C_r^{'}-c_y \end{bmatrix} \tag{2}$$

where a=25.39 , b=14.03 , ec_x=1.60 , ec_y=2.41 ,θ=2.53 , c_x=109.38 and ec_y=152.02 are computed in the $YC_b^{'}C_r^{'}$ space.

III. SUPPORT VECTOR MACHINES

Support Vector Machines theory [10] was developed by Vapnik et al and is a new statistical pattern classification method which bases on VC-dimensional theory and the idea of structural risk minimization. The statistical learning theory establishes a good theory framework for the machine learning problems of small samples. It can solve the small samples, non-linear, high-dimension, local minimum point, and other practical issues. Its core idea is to make learning machines and limited training samples adapt to each other. Using quadratic programming optimization, we can get the global optimal solution, and solve the local minimum problem which the artificial neural network can not avoid. Due to adopt kernel function, it can subtly solve the dimension problem, make the algorithm complexity has nothing to do with the sample dimensions, and is well suited to deal with non-linear problem. Moreover, SVM adopt the structural risk minimization principles, so it has very good generalization ability.

Given a binary classification with ℓ sample points: (x_i,y_i), i=1, …, ℓ, where $x_i \in R^n$ and $y_i=\{+1,-1\}$ is the classify label. There is an optimal hyperplanes which has the maximum margin of separation between the classes. And the margin equals $2/\| w \|$. Minimizing the magnitude of the weights can maximize the margin. The problem is then transformed into a quadratic programming problem which can be solved by SVM with using Lagrange optimal algorithm.

$$\begin{aligned} &\max Q(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathrm{x_i}\cdot \mathrm{x_j}) \\ &subject \quad to \quad \sum_{i=1}^{l} y_i\alpha_i=0 \quad 0\le\alpha_i\le C \quad ;i=1,\cdots,l \end{aligned} \tag{3}$$

Notice that the positive constant C is a regularization factor, and it controls compromise between margin and misclassification error. $K(\mathrm{x_i},\mathrm{x_j})$ is kernel function. A kernel $K(\mathrm{x_i},\mathrm{x_j})$ is called an SV kernel if it satisfies the Mercer theorem. $K(\mathrm{x_i},\mathrm{x_j}) = <\mathrm{x_i},\mathrm{x_j}>$ is LSVM kernel function. For non-linear SVM, polynomial kernel, Gaussian RBF, sigmoid kernel etc are commonly used in kernel function.

If α_i^* is the optimal solution, w^* is the weight vector of the optimal hyperplane.

$$w^*=\sum_{i=1}^{l}\alpha_i^* y_i \mathrm{x_i} \tag{4}$$

w^* is the linear combination of training sample vector. The training samples are called support vectors when α_i^* is not equal to zero.

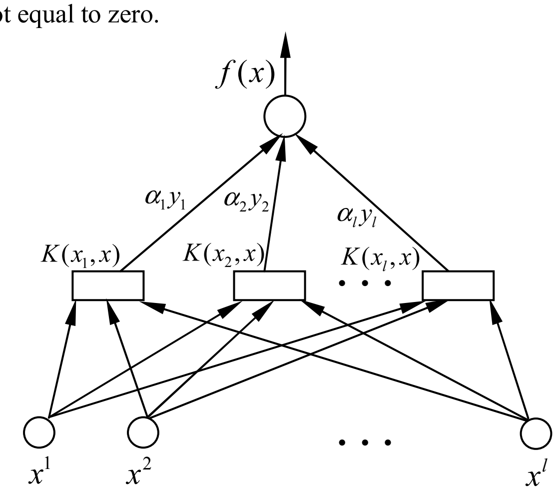

Figure 1. Support Vector Machines

SVM can be depicted in Figure 1, and the decision function can be written as

$$f(\mathrm{x})=\mathrm{sgn}\{(\mathrm{w}^*\cdot \mathrm{x})+\mathrm{b}\}=\mathrm{sgn}\{\sum_{i=1}^{l}\alpha_i^* y_i K(\mathrm{x_i}\cdot \mathrm{x})+\mathrm{b}^*\} \tag{5}$$

IV. FACE DETECTION ALGORITHM BASED ON FACIAL FEATURES AND LINEAR SUPPORT VECTOR MACHINES

The process of face detection algorithm is described in Figure 2. At first, the sample images and testing images are pre-processed with lighting compensation, denoising, etc. After nonlinear color transformation, the skin color candidate regions can be detected and segmented from the background. The samples are collected from the skin color regions of original image. After applying the wavelet transform and discrete cosine transform, the sample feature vectors are put into the SVM classifiers for training or classifying. Finally, the eyes and mouth detection are carried out to finalize the face regions.

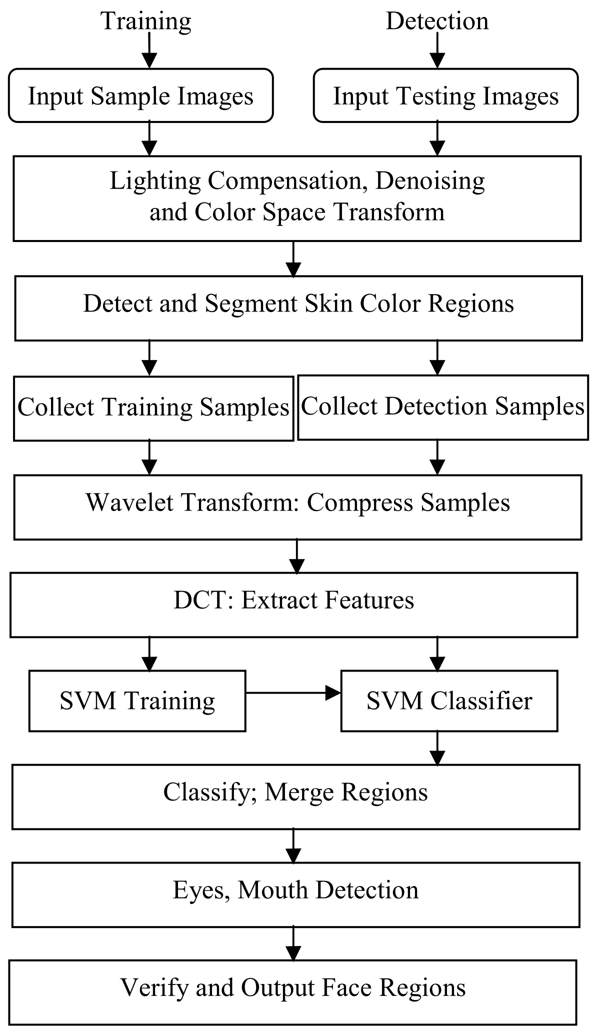

Figure 2. Face detection algorithm flow chart

A. Skin Color Detection

At first, the example images and testing images are received treatment with lighting compensation technique, and are filtered with a 3×3 low-pass filter to reduce the effect of noise. If a pixel meets (1) in the $YC_b'C_r'$ color space, then it is considered to be a skin color pixel and is marked as 1. Otherwise, it is considered to be a non-skin color pixel and is marked as 0. As a result, we get a binary image. After morphology processing, the skin color regions can be detected accurately.

B. Sample Training

1) Sample Selection Scope

In the traditional methods, face samples are always selected directly from the database such as CBCL and non-face samples are selected from the database or the gray scenery images, such as trees, plants, buildings, and so on. However, we select samples from color images. Moreover, non-face samples are selected from the skin color regions. So its selecting scope is narrowed.

2) Collecting Training Samples

A total of 300 sample images are collected from the Internet, photos and so on. After processing with lighting compensation, denoising，detecting skin color regions, and so on, we manually extract 1,000 face samples and 1,300 non-face samples. And the sample size is 20×20 pixels. In order to improve generalization ability and increase robustness to a little change of face size and orientation, the samples are processed as follows：(1) Rotated a certain angle clockwise or counterclockwise. (2) Mirrored along the Y axis. (3) Scaled. After the above processing, we get more than 12,000 face samples and 15,600 non-face samples.

3) Extracting sample features

To reduce the SVM dimension and the training time, sample processing is done as follows: (1) The sample is compressed by applying the wavelet transform and selecting the low-frequency part. (2) A mask is used to remove the four corners' background interference, and reduce the feature vector dimension. (3) The discrete cosine transform has a good feature extraction and excellent data compression and have less computing features. Moreover, in order to further increase robust to lighting, the coefficients of the triple color components are discriminatively selected. Y component is selected anterior 50 coefficients, and Cb and Cr component is selected anterior 60 coefficients. As a result, the total number of coefficient adds up to 170. So the dimension of sample feature vector is reduced from 1200 to 170. The dimension of SVM is effectively reduced.

4) Classifier Training

As the face region has rich texture, eyes, mouth and other information on it, there is a large difference between the face and skin-likelihood background or other body regions, such as the neck, shoulders, and arms. The LSVM kernel function $K(\mathrm{x_i},\mathrm{x_j}) = <\mathrm{x_i},\mathrm{x_j}>$ is adopted. And a LSVM classifier is designed to classify and is used LibSVM [11] to training the samples. The positive constant C is equal to 200.

C. Face Detection

1) Selecting and Classifying the Detection Samples

As the number and size of faces are often different in images, in order to detect all the faces, we have to take pyramid analysis. Every image will be lessened by 1.2 times until the size of the skin color region is 20×20 pixels. In every scale, we scan the original color image with a windows size of 20×20 pixels from left to right, top to bottom in the effective skin color regions, and intercept image as a detected sample. After the detection sample is processed with a mask, the wavelet transform, and discrete cosine transform, the sample feature vector is put into the SVM classifiers to classify.

In the process of selecting sample, we notice the following: (1) Because SVM use 20×20 image for the training, and because the region is less than 20×20 pixels, the facial features will not provide sufficient detail for face recognition and other application, the skin color region rectangle which height or width is less than 20 pixels is considered non-face region. (2) According to the skin color pixels in the candidate region should be a certain proportion. The experiment indicates that the region whose skin color pixels are less than 35% will be non-face region.

2) Merging the Overlap Regions

In the general, the same face will be detected very many times in the adjacent position or different scales. We need to merge the overlap regions, and obtain the region's location and size. When the candidate region overlaps more than 70%, we believe that the face candidates represent the same face. Then we merge them. According to the average of the location and size, we determine the location and size of the face candidate regions. In addition, the false detections often are the regions which are repeatedly detected very seldom.

D. Verifying the Face Regions

Eyes and mouth is the most distinct facial features. The color component is great different between skin and eyes and mouth in the color space, it is easy to detect the eyes and mouth candidate. Finally, we can easily verify the face regions by the combination eyes and mouth of the geometric relationship.

1) Exclude Most Non-eyes and Non-mouth Regions

In the face candidate region, according to the geometric features of connective region, such as the region's size, scale of length and width, location and so on, we can remove most the areas which are not entirely to be the eyes or mouth. For example, we can remove the regions that are much smaller and larger about 3% of the entire face candidate region.

2) Eyes Detection

As the iris is greater different from the skin, it is obvious that eyes are non-skin color regions. In the YCbCr color space, the Cb and Cr component of eyes and skin contains bigger difference. Moreover, the gray value of the eyes is Very small. Y component is in the range from 0 to 120. In addition, the Cb is very higher than the Cr in the eyes region. Banding together eyes' chrominance and luminance [9], we can accurately detect the location and size of the eyes.

3) Mouth Detection

Mouth is obviously non-skin color region too. Compared with other organs, the mouth's Cr is much higher, and the mouth's Cb much lower. Increasing the difference between Cb and Cr [9], we can also accurately detect the location and size of the mouth.

4) Verifying the Face Regions

The center of the eyes and mouth can structure a facial triangular model. According to the geometric relationship between the eyes and mouth, such as geometry and orientation of the triangle, we can draw a conclusion about verifying the face regions.

V. EXPERIMENTAL RESULTS

To evaluate the efficiency of algorithm proposed in this paper, we test several algorithms on 115 color images, which were collected from photos and the Internet. These images were taken under different lighting conditions and complex backgrounds, and their resolution ranges from 320×240 pixels to 1024×768 pixels. These images contain one face or several faces per image. There are 692 faces in total. The faces are different in size, location and facial expression. The experiments were carried on the P4 computer which has 1.83GHz dual-core CPU and 1GB memory, using Visual C++6.0 for programming. Table Ⅰ shows the test results with different algorithms.

TABLE I. TEST RESULTS

Detection Algorithms	Correct Detection	Missing Detection	False Detection	Correct Rate
Based Skin Color and Facial Features	639	53	265	92.3%
SVM Proposed in [7]	529	163	219	76.4%
Hierarchical SVM	617	75	176	89.2%
Algorithms Proposed in This Paper	655	37	83	94.7%

Compared with the algorithms shown in Table Ⅰ, we can see that our algorithm can promote detection correct rate and less false detections. The experimental process also demonstrates that our algorithm can speed up face detection.

Figure 3 shows some detection results obtained by our algorithm.

Figure 3. Some experimental results

VI. CONCLUSION

This paper proposes a face detection algorithm based on facial features and LSVM. The statistical learning method is strongly related to the training samples. With intentionally selecting samples and detection regions from the skin color regions, the effectiveness of samples and the performance of classifier are improved. Moreover, the sample images are compressed by wavelet transform, and sample feature vector is extracted and reduced by discrete cosine transform, the training difficulty of SVM classifier and the matching times are obviously reduced and the detection is speeded up. At last, the face candidate regions are verified by detecting eyes and mouth. Experiments show that algorithm proposed in this paper achieves higher detection accuracy, lowers false detection rate, greatly speed-up detection and is highly robust to lighting.

REFERENCES

[1] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey" [J], IEEE Trans. Pattern Analysis and Machine Intelligence, 2002, 24(I).

[2] Hjelm and B. K. Low, "Face Detection: A survey" [J], Computer Vision and Image Understanding, 2001, 83(110.3), pp. 236-274.

[3] Yang J, Lu W, Waibel, "A Skin Color Modeling and Adaptation"[R], Technical Report CMU2CS2972146, CMU, Pittsburgh, PA, 1997.

[4] Chen Duan-Sheng, Liu Zheng-Kai1. "A Survey of Skin Color Detection" [J]. Chinese Journal of Computers, 2006, 29(2), pp. 194-207.

[5] Vapnik V N. "The Nature of Statistical Learning Theory" [M], New York: Springer-Verlag, 1995.

[6] C. Campbell. Algorithmic Approaches to Training Support Vector Machines: a Survey[J]. Proceedings of ESANN2000, 2000: pp. 27-36.

[7] Osuna E, Freund R. "Training Support Vector Machines: an Application to Face Detection" [A]. Proc. Of Computer Vision and Pattern Recognition [C].San Juan, Puerto Rico, IEEE Computer Soc,1997: pp.130-136.

[8] Ma Yong, Ding Xiaoqing. "Face Diction Based on Hierarchical Support Vector Machines" [J], J Tsinghua Univ (Sci & Tech), 2003,43(1), pp.35-38.

[9] Rein-Lien Hsu and Anil K. Jain. Face Detection in Color Images[C]// IEEE Trans. Pattern Analysis and Machine Intelligence, 2002,

[10] Zhang Xuegong, "Introduction to Statistical Learning Theory and Support Vector Machines" [J]. Acta Automatica Sinica, 2000, 26(1):pp.32-42

[11] C. C. Chang, C.J. Lin. LIBSVM -- A Library for Support Vector Machines：http://www.csie.ntu.edu.tw/~cjlin/libsvm/ 2008