A Robust & Fast Face Detection System

Ritu Verma, Anupam Agrawal and Shanu Sharma
Indian Institute of Information Technology Allahabad, INDIA
Email: rituverma1021@gmail.com
Indian Institute of Information Technology Allahabad, INDIA
Email: {anupam@iiita.ac.in, shanu.sharma1611@gmail.com}

Abstract: Human face detection is a significant problem of image processing and is usually a first step for face recognition and visual surveillance. This paper presents the details of face detection approach that is implemented to achieve accurate face detection in group color images which are based on facial feature and Support Vector Machine. In the first step, the proposed approach quickly separates skin color regions from the background and from non-skin color regions using YCbCr color space transformation. After the detection of skin regions, the images are processed with wavelet transforms (WT) and discrete cosine transforms (DCT) as a result of which the 30×30 pixel sub images are found. These sub images are then assigned to SVM classifier as an input. The SVM is used to classify non-face regions from the remaining regions more accurately, that are obtained from previous steps and having big difference between faces regions and non-faces regions. The experimental results on different types of group color images show that this approach improves the detection speed and minimizes the false detection rate in less time and detects faces in different color images.

Index Terms: Face Detection; Skin Color Detection; Wavelet Transform; Discrete Cosine Transform; Support Vector Machine.

I. INTRODUCTION

A face detection system is a system that determines the locations and sizes of human faces in arbitrary (digital) images. It detects facial features from images and ignores all other things, like buildings, trees etc. Recently, researchers have proposed to detect face by method combining features and color to obtain a high performance and high speed results [1], [4] and [13]. Detecting faces is a crucial step in the identification applications for example airport security, law enforcement etc. Most of the face recognition and face tracking algorithms assumes that the initial face localization is known. The main merit of any good approach is to provide fast and high detection ratio and can deal with faces in complex background.

In this paper, implementation of a robust face detection algorithm which is based on facial feature and LSVM (linear support vector machine) is presented. This algorithm deals with different complexities and provides high speed and high detection ratio. Different complexities include finding number of faces in group image, varying illumination, occlusion and complex background present in an input image.

The skin color is a significant feature of a face. It has a strong cluster feature of YCbCr and HIS color space [1]. In YCbCr, Y stands for the “luma” (luminance) which is brightness. Cb and Cr stand for the “color difference” of blue – luma (B-Y), and red – luma (R-Y) respectively. Skin color gives more reliability because it is not affected by body posture and facial expression. It is easily distinguished from the background color. Hence the face detection approaches, based on the skin color, are widely used. But it is not sufficient to absolutely and precisely detect the face only by using skin color information. When several faces are very near to each other or the face regions and other body regions are close or skin-likelihood background is connected together to the face, it often increases the false detection ratio. This problem can be handled by detecting the false candidate regions with statistical methods. In this face detection system the sub images of faces are very small in size for which the statistical learning is used. Statistical learning theory is currently the best theory for small samples statistics estimates and projection learning. SVM theory is established on the basis of statistical learning theory; its objective is to resolve the problem of classification of small samples.

The outline of the paper is prepared as follows: The summary of literature survey described which is similar to my system and few face detection methods with their merits and demerits. Section III explains the details of the implementation and methods we have been used. In section IV the results of this face detection approach on various types of images are discussed and in section V the conclusion and scope for the future work are explained.

II. RELATED WORK

Face detection technique is an open challenge from last many years, and various solutions addressing the face detection problem have been proposed under different categories which are discussed below. Face detection is not an easy method as the detection is affected by many internal and external factors. Few main Face Detection Methods are as follows:

A. Knowledge-Based Method:

In this method the relationship between facial features of test image is used to represent the content of the face and then encode picture digitally as a set of rules and to reach the finest scale. It is a top down approach [5]. Merits and demerits of knowledge-based method are as follows:

Merits
• It is simple to describe the features of face and their relationship by using simple rules.
• By coded rules first facial features of image are extracted then candidate faces are identified.
Demerits
- Translation of human knowledge into precise rule is very difficult.
- General rules may find many false positives.

B. Template Matching Method:
This method is based on finding the co-relation between a test sub image and the pre-defined stored face patterns. The predefined images might be the whole face or individual face features such as nose, eyes, mouth, eyebrows, and lips [5]. Algorithms used under this method are:

Predefined Face Templates:
In predefined face templates several templates for the whole/individual or both parts (whole & individual) of the face are stored.

Deformable Templates:
In this an elastic facial feature model as a reference model is stored and the deformable template mode of the object of interest is fitted in.

Merits and Demerits of Template Matching Method are as follows:

Merits
- It’s simple and easy to implement.

Demerits
- Templates have to be initialized near the face images.
- Difficult to enumerate templates for different poses.

C. Feature-Invariant Approach:
In this approach faces structural features are not changed under different conditions, such as varying viewpoints of cameras, pose angles, and/or illumination conditions.

Algorithms used under this approach are:

Colour-Based Approach:
Colour based is also called skin-model based method. This approach is based on the fact that different skins from different races are clustered in a single region and makes use of the skin colour as indication to the presence of human beings [1], [4] and [6].

Facial-Feature Based Approach:
In this method global and/or detailed features are used for face detection. It has become popular in recent days. The global features (e.g. skin, size and shape) are firstly used to detect the candidate area after that they are tested using detailed features (e.g. eyes, nose, and lips) [13].

Merits and Demerits of Feature-invariant approach are as follows:

Merits
- Features are invariant in different poses and orientations of the faces.

Demerits
- Difficult to locate facial feature due to various complexities (illumination, occlusion etc.) in an image.
- Difficult to detect features in complex background.

D. Appearance-Based Method:
This method learns the templates from the set of training images. It finds the relevant characteristics of face and non-face by using statistical analysis and machine learning techniques [3] and [7]. Algorithms used under this method are:

Eigen Faces:
These are also called the eigenvectors, in which different algorithms are used to approximate the eigenvectors of the auto correlation matrix of a candidate image [19].

Neural Network:
A network of neurons (simple element) called nodes used is to perform function in parallel. Central nervous system gave this idea of neural network. These networks are trained for the detection of faces by providing it, face and non-face samples [15].

Support Vector Machine:
Support vector machine are learning machine and it makes binary classification. The idea is to enlarge the difference or margin between the vectors of negative and positive sets and obtain an optimal boundary which separates two sets of vectors [8] and [14].

Hidden Markov Model:
It is also abbreviated as HMM model and can be considered as simple dynamic Bayesian network. Hidden Markov Model is a class of statistical model which uses the statistical properties of a signal that model the processed system. The Markov parameters should be taken from the observed parameters [16].

Merits and demerits of Appearance-based method are as follows:

Merits
- Use powerful machine learning algorithms and it has demonstrated good empirical results.
- It offers to detect faces in various poses and orientations.

Demerits
- It is usually needed to look for the space and scale.
- It requires lots of positive and negative examples.

II. DETAILS OF THE APPROACH IMPLEMENTED

The flow chart of a proposed approach is shown in figure1.
Steps for Face Detection:

1. First give a RGB image as an input image to the Skin color model.
2. The Skin color model converts the RGB image to the YCbCr color space model [18].
3. For handling varying lighting conditions convert this output image in YC’bC’r color space by the elliptical formula [7].
4. For reducing noise effects filter this image by 3×3 low pass filter, and then apply morphology (dilation) operation to get a binary image [18].
5. Find the skin regions based on above binary image.
6. The discrete wavelet transform (DWT) decomposes the given input image into a set of sub-bands of different resolutions and selects the low frequency parts. The new generated top left low frequency sub-bands are nearly equal to the original image [18].
7. Take the output of the DWT to the DCT and use 30×30 size window to pick up the significant information of signal energy [11].
8. Support Vector Machine is used for classification to construct an optimal hyper-plane which has a maximum margin of the separation between the face and non-face classes [8]. We have taken 30×30 size of windows as an input and separate these in faces or non-faces by the classification.
9. Obtain the final face detected output image.

Details of main components of the approach are given below:

A. Skin Color Model And Segmentation:

In order to apply this method in the real time system, skin color detection is adopted; de-noising and lighting compensation are the initial steps of skin color model. This is because the lighting condition and noise has great effect on the skin color detection. YCbCr color space transformation is faster than the other approaches and popularly used in skin color detection [2]. YCbCr color space is developed for television systems, and it is luminance separated color space so it is widely used in mpeg, jpeg and other video compression standards.

First linear conversion of RGB color space to YCbCr color space is obtained, but for further reduction in the lighting effect and to obtain a good result of skin color cluster, a segmented non-linear conversion algorithm [7] is used which converts YCbCr color space into the YC’bC’r color space. Segmented skin color regions are obtained by the elliptical cluster method for the skin tones in the transformed YC’bC’r space. It is described in equations (1) and (2) as given below [7].

\[
\left(\frac{x - ec_x}{a}\right)^2 + \left(\frac{y - ec_y}{b}\right)^2 < \Theta \quad \cdots \text{(1)}
\]

\[
\left[\begin{array}{ccc}
\sin \Theta & \sin \Theta & 0 \\
\cos \Theta & \cos \Theta & 0 \\
0 & 0 & 1
\end{array}\right] \left[\begin{array}{c}
C_r \\
C_b \\
C_y
\end{array}\right] = \left[\begin{array}{c}
c_x \\
c_y
\end{array}\right] \quad \cdots \text{(2)}
\]

Where \(a = 25.39, b = 14.03, ec_x = 1.60, ec_y = 2.41, \Theta = 2.53, c_x = 109.38, \) and \(c_y = 152.02\) are computed in the YC’bC’r space [7].

The images are received after lighting compensation technique, and are filtered with a 3×3 low pass filter [18] which is used for minimizing the effect of noise. If then the pixel satisfies equation (1) in elliptical cluster method (YC’bC’r color space), it is marked as 1 and has to be considered as skin color pixel. Otherwise, it is marked 0 and has to be considered as non-skin color pixel. It provides an output binary image after the above process. Finally it can detect skin color regions accurately after morphological (dilation) operation [18].

B. Discrete Wavelet Transform:

For reducing the training time and SVM dimension, the samples are compressed by wavelet transform (WT). Here using the discrete wavelet transforms which is based on sub-band coding and it is found to create a fast computation of WT [12]. It is easy to execute and minimize the computational time and resources required.

The discrete wavelet transform decomposes the input frame of image into a set of sub band of different resolutions. The new generated sub-band is nearly equal to the original frame. DWT is a time-scaled representation of the digital signal and is found by digital filtering techniques [18]. The amount of the information present in the signal is measured and this is termed as the resolution of the signal which is to be finding out by several filtering operations and it is given by up-sampling and down-sampling phenomena. The dilation function of discrete wavelet transform is represented by a tree of low & high pass filters. Low pass filters are transforming in each step. The original signals are continuously decomposed into the subpart of lower resolution and the high frequency components are not analyzed.

Wavelet coefficients are created into wavelet blocks in which horizontal, vertical and diagonal edges are the sub images of real image, it is shown in figure2. The upper most left sub image represents the superior level of low pass sub image. The concept of wavelet block gives an association between coefficient and what they represent spatially in the frames [10].

Figure2: wavelet block are reconstruction of wavelet coefficient.

This is a four level discrete wavelet transform [10].
C. Collecting Training Sample:

In the previous methods training samples are collect from the database directly and the non-face samples are selected from the scenery images, such as building, plants, trees and so on. So that it narrows the selecting scope. But here the training samples are selected after the processing with color transform, de-noising, and detection of skin color regions and so on. Here we use 12 images for testing purpose which are collecting from personal digital camera and also from the database [17]. After the initial steps like color space transformation, lighting compensation and detection of skin regions we get scaled images. From the scaled images we extract 30×30 pixel sub-images and here we get around 700 sub-images from 12 testing images and extract them in 150 faces and 550 non-faces.

D. Discrete Cosine Transform:

The DCT is a good example of the transform coding [18]. The recent JPEG standard images use the DCT as its basis. The discrete cosine transform relocates the high valued energies (information) to the upper left corner of the image and the lesser energies are relocated in other areas [11]. Discrete cosine transform is a unique method that has near-optimal energy compaction property [9]. It separates the given image into sub-bands (parts of image) on the basis of visual quality. The DCT has a great feature extraction and excellent data compression and has less computing features. It gives robustness for detection in lighting effects or variations.

Energy Compaction is the main property of DCT [11]. Having a power to produce a transformation scheme can be directly approximated by its ability to compact input data into a few possible coefficients. It allows quantizer to remove coefficient with relatively small amplitudes and reconstruct image without any visual distortion. DCT exhibits excellent energy compaction for highly correlation sub-images. In the transform coding, the pixels in an image displays a certain level of correlation with neighboring pixels. Same problem is there in video transmission which shows very high correlation of adjacent pixels in consecutive frames. We take the output of Discrete Wavelet Transform as an input to the Discrete Cosine Transform and use 30×30 size window to pick out the significant information of signal energy. The sample feature vector is extracted and compacted by DCT [7].

E. Support Vector Machine:

A SVM is a supervised learning technique form of machine learning, and it is applicable for classification and regression. This support vector machine theory is developed by Vladimir Vapnik & his team in 1995 at AT&T Bell Laboratories, and the principle is based on structural risk minimization, so it has very good generalization ability [8]. Generalization means the summation of data and knowledge.

The main aim of statistical learning theory is to present a framework for studying the problem of inference, which is of gaining knowledge, making predictions, making decisions or constructing models from a set of data. The proposed method adopts a kernel function so it is able to solve the dimension problem, and is well suited for non-linear problem. A LSVM classifier is designed to classify and used LibSVM [8] to train the samples. The LSVM kernel function is adopted here-

\[ K(x_i, x_j) = \langle x_i, x_j \rangle \]  

(3)

In a binary classification with / sample points:

\[ (x_i, y_i) = 1, 2, 3 \ldots l \]  

(4)

Where \( x_i \in R^d \) and \( y_i = \{+1, -1\} \) are the classifying label [7]. This system finds faces by thoroughly scanning an image for face like patterns at several possible scales, by isolating the original image into overlap sub-images and determines them into appropriate class face or non-face by using support vector machine. The figure 3 shows the geometrical interpretation of the technique support vector machine provides in the framework of the face detection. The vital use of support vector machine is in the classification step, which is the essential part of the work.

By using support vector machine classify all window patterns and if the class matches a face then make a square around the face in the output image.

IV. EXPERIMENTAL RESULTS

Here evaluation of proposed methodology on a face image database, and construction of the database for face detection from personal photo collections and internet [17] is done. These color images or the database has been taken under different complexities, like detecting possible faces under varying illumination conditions and occlusion in group photographs with complex backgrounds. With high detection rate of 87.65% accuracy, this approach can detect all possible faces in between range (9.38sec to 11.97sec) of time. The face detection time depends on the complexities of the testing color images. Further the discussed approach is able to detect multiple numbers of faces with broad range of facial variations in an image.

A. Discussion for the output images shown in section B are given below:

1. The first input image is the original RGB image which we get either from the personal dataset or from the internet datasets [17], having different complexities. For example the given input image1 has varying illumination over different faces and has complex background.
2. Perform low pass filtering to reduce effect of noise and for handling varying lighting condition use elliptical formula (as discussed in above) on the input image. From this we get the binary skin map image.
3. Third image shows the skin region detected image of the input RGB image. Here we separate the background of image from the skin color regions.

4. For the fourth image, perform the dilation operation (morphological operation) on the 2nd skin map region image. The dilation operation which accepts the structuring element objects, known as STRELs [18].

5. The fifth image shows the dilated skin region detected image of the input image after applying the above operations on the 4th image.

6. Apply discrete wavelet transform to get a sixth scaled image.

7. After getting the scaled image apply discrete cosine transform. By applying this process the image is divided into the 30×30 sub-images, and we train all sub-images as a face or non-face sub-image.

8. In seventh image, Support Vector Machine (SVM) is used for classification of data to construct an optimal hyper-plane which has a maximum margin of the separation between the face and non-face classes.

9. Finally we obtain the final face detected output image (image8) after classification, where faces are enclosed in boxes around them.

Here, we have collected 12 testing color images of different sizes and different complexities. In these 12 testing group color images, first six images (1 to 6) are taken from personal digital camera and the next six images (7 to 12) are taken from the face detection datasets “Bao Face Database” [17]. Total 81 faces are there in 12 images in which 71 faces are detected successfully. This approach gives accuracy 87.65% with a good speed. After the training time of the faces and non-faces it can able to detect the possible faces in between range 9.38sec to 11.97sec. Its detection timing depends on the complexities of the images. Table1 and Table2 show the results of finding faces in different given input images.

**TABLE I:**

<table>
<thead>
<tr>
<th>Sr. no.</th>
<th>Number of faces in images</th>
<th>Correct detection of faces</th>
<th>Missing detection of faces</th>
<th>Detection time of faces(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>9.87</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>10.16</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>9.77</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>9.64</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>9.38</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>11.97</td>
</tr>
</tbody>
</table>

**TABLE II:**

<table>
<thead>
<tr>
<th>Sr. no.</th>
<th>Number of faces in images</th>
<th>Correct detection of faces</th>
<th>Missing detection of faces</th>
<th>Detection time of faces(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>12</td>
<td>8</td>
<td>4</td>
<td>10.24</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>9.99</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>9.96</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>10.88</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>11.44</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>11.20</td>
</tr>
</tbody>
</table>

Complexities in different input images which are shown in below section B and section C are:

1. Image1 has complexity of varying illumination over different faces and has complex background (skin likelihood background).

2. Image9 has complexity of occlusion and has complex background.

3. Image10 has complexity of tilted faces.

**B. The output images (2 to 8) generated by various steps on input image (1) are given below:**

- Image1. The original RGB image
- Image2. Skin map image
- Image3. Skin region detected image
- Image4. Dilated skin map image
- Image5. Dilated skin region detected image
- Image6. Scaled image after applying DWT image
- Image7. Classification by SVM image
- Image8. Final face detected image
C. The output for more images with different complexities:

![Image9. Face detected image](image9.png)  ![Image10. Face detected image](image10.png)

V. CONCLUSION AND FUTURE WORK

This paper discusses a robust & fast face detection approach and its implementation is based on facial feature and LibSVM. The statistical learning theory is related to the training samples. Selected samples and regions which are found from the skin color regions by non-linear conversion are used; the strength of samples and the functioning or the performance of classifier is improved. For the compression purpose we use here discrete wavelet transform and for extracting the feature vector of sample images we use discrete cosine transform, so the resultant matching time and the training difficulty of support vector machine are obviously reduced and there is speeding up the algorithm. Result shows that the algorithm achieves good (around 87.65%) detection accuracy, lower false detection rate and improved speed, which makes the algorithm highly robust.

Further the present work may be extended to reduce the false detection rate, solve the problem of shifted boxes and improve its accuracy for face recognition.

ACKNOWLEDGEMENT

The authors would like to express sincere gratitude to the Director of the Institution Dr. M.D Tiwari for providing excellent computational facilities and stimulating work environment for carrying out the research work.

REFERENCES