A Dynamic Human Face Detection Algorithm Based on Video Sequence

Yanfen GAN, Chunyuan FENG
Information science and technology department
Guangdong University of Foreign Studies
South China Business College,
Guangzhou, 510545, China
Tel.: +8613751800978
E-mail: Yanfengan1981@Gmail.com

Received: 17 September 2013   /Accepted: 22 November 2013   /Published: 30 December 2013

Abstract: In the security monitoring of banks and confidential departments, it is essential to monitor and record unusual entrants through detection of dynamic human faces. In this paper, a dynamic human face detection algorithm based on video sequence is proposed. The new algorithm consists of three parts: detection of moving objects, approximate skin color detection and frontal view human face detection. The algorithm first quickly detects moving objects through a method based on color curve proportion difference, then based on which dynamic human faces are detected through approximate skin color segmentation and structure-based second identification of frontal view human faces. Test results have shown that the dynamic human face detection algorithm based on video sequence proposed in this paper is effective. Copyright © 2013 IFSA.

Keywords: Moving object detection, Face detection, Skin color detection, Video sequence.

1. Introduction

Nowadays, some important organizations in our society such as bank vaults and some departments where confidential documentation is kept often suffer from theft and destruction caused by criminals, thus resulting in serious losses. How to provide enhanced safety guarantee under such environmental conditions has become a problem to be solved. Dynamic human face detection is crucial to the safety monitoring over those departments. Song Hong, Shi Feng, et al. [1, 2] proposed a quick and automatic multiface detection algorithm based on the multi-source information fusion technology and knowledge about human face structure in the context of the intelligent video monitoring system for human face detection. This algorithm integrates the moving information and skin color information in images and obtains candidate face regions by using relevant prior knowledge. Niu Deji, Zhan Yongzhao and Song Shunlin [3] proposed a method for real-time face detection and tracking in video sequences in collaborative systems. This method establishes the skin color model of human faces using multiple training samples based on the color distribution characteristics of user-specific objects (such as human faces). This model and human face characteristics are then used to segment and match the color images so as to determine human faces from the candidate regions.

From the above references, we know there are several challenges for dynamic face detection base on structural features and video sequence. First, in detection of moving objects, the difference image method is frequently used in video monitoring systems to detect moving objects. However, when the difference image method is used to detect objects, the
illumination changes will affect the detection results \[4, 5\]. Second, skin color detection under different illumination conditions will encounter problems \[6\]. Finally, in the second identification of human faces, the object with similar structure to human faces will prevent us from finding out correct human faces.

To solve the above problems, this paper proposes the color curve proportion difference algorithm which can be used to quickly and accurately detect moving objects while avoiding inaccurate detection resulting from illumination changes. In approximate skin color detection, a great number of data are collected to set up the skin color model. Through integration of the two color spaces YIQ and YUV, a strict threshold range is selected to enable the skin color of human faces to be better screened. In the second identification of human faces, this paper combines a number of methods \[8-10\], including image mosaic, horizontal line detection, and extraction of characteristics and use of structural rules of human faces, which can accurately locate human faces and distinguish human faces from non-human faces. Fig. 1 shows the dynamic face detection algorithm.

The reminder of this paper is organized as follows: Section 2 introduces the quick detection algorithm for moving objects. The approximate skin color detection and frontal view face detection are presented in Section 3 and 4. Experimental results and conclusion are described in Section 5 and 6 respectively.

2. The Quick Detection Algorithm for Moving Objects

In this part, the first step is to determine whether there are moving objects in adjacent frames. If no changes take place, replace the background; otherwise, compare the color single-channel value with the gray value at each point according to the change difference of the color single-channel curve between the previous frame and the next frame and do accumulating projections of the ratios. After that, the projection is further processed to obtain the positioning map.

2.1. Color Curve Proportion Projection Difference Algorithm

The color curve proportion projection difference algorithm first judges in real-time whether there exist moving objects in the previous and the next frames with an interval of \(\Delta t\) based on (4) in this paper. If no moving objects exist, immediately replace the previous frame with the next frame as the background. When moving objects appear, the single-channel color projection of the \(k\)th and \((k+1)\)th frames is shown as Fig. 2.

\[
U_{(x,y,j)} = \sum_{i=1}^{m} U_{(x,i,j)},
\]

where \(x=1,2\ldots m, j\) stands for the \(j\)th frame.

We use the figure size of \(m \times n\), \(m = 700, n = 525\). As we see, \(U_{(x,y,j)}\) is the value of any channel in the three primary colors at the \(j\)th frame \((x, y)\) of the color image. This paper selects the red channel. \(U_{(x,0)}\) is the accumulating projection in the vertical direction of \(U_{(x,y,0)}\). According to (2), it can be obtained that the light colored solid line in Fig. 3(a) is the accumulating projection of the \(k\)th frame. The dark astroid \(U_{(x,k+1)}\) is the accumulating projection of the \((k+1)\)th frame. The figure shows that there are equal projected intervals in some parts of the two curves, which is caused by gradual illumination changes. If subtraction is made directly at this time, the location will be inaccurate.

In order to eliminate this influence and highlight unequal intervals, this paper adopts the idea of accumulation of proportion differences here. In the
same frame, when the single-channel value $U(x,y,j)$ is affected by the gradual illumination changes, the trend of the gray value $V(x,y,j)$ is the same, and their ratio $U(x,j)$ is slightly affected by illumination. It is described as follows:

$$U(x,j) = \sum_{j=1}^{m} \left( \frac{U(x,j)}{V(x,j)} \right)$$

where $V(x,y,j)$ stands for the gray value at the $j$th frame $(x, y)$.

It can be obtained from (2) that the light colored solid line $U(x,k)$ in Fig. 2(b) is the proportion accumulating projection of the $k$th frame and the dark astroid $U(x,k+1)$ is the proportion accumulating projection of the $k+1$th frame. The figure shows that accept the part where moving objects appear (between 300 and 400), other parts are basically consistent. Finally, according to the difference $U(x,d)$ between curve $U(x,k)$ and curve $U(x,k+1)$ (as shown in Fig. 3(b)), the abscissa of the moving object can be determined.

$$U(x,d) = |U(x,k) - U(x,k+1)| \left| \sum_{j=1}^{m} \left( \frac{U(x,j)}{V(x,j)} \right) - \sum_{j=1}^{m} \left( \frac{U(x,j+1)}{V(x,j+1)} \right) \right|$$

The result is shown in Fig. 4(a), from which it can be known that there will be obvious crests where the moving object appear. Therefore, the moving object can be located according to the crests.

As the proportion projection difference map contains some high-frequency noises which will affect the detection, in order to better distinguish “crests” and “troughs”, Gaussian smoothing is applied to the images. Here, filtering is done using the two-dimensional Gaussian function $G(r)$.

### 2.2. Moving Objects Determination and Exist and the Location

When moving objects appear, the gap between two frames will increase quickly. The condition to determine whether it is necessary to replace the background is shown in (4).

$$U_{1} = U_{1}^* - \sum_{j=1}^{m} U_{1}^* / m$$

$$Y = \begin{cases} 1, \text{while } \max(U_{1}^*) \leq T \\ 0, \text{while } \max(U_{1}^*) > T \end{cases}$$

Fig. 2 (a). Background image without a moving object and the background image with a moving object.

Fig. 2 (b). Background image without a moving object and the background image with a moving object.

Fig. 3. two single-color projection difference curves and two projection proportion differential curves.
According to the experiment, $T$ is assigned 15. $Y = 1$ means that no moving objects appear and the background can be replaced. $Y = 0$ means that a moving object appears and it is essential to locate it timely. The waveform of $U_{10\theta 0}$ is shown in Fig. 5(a). There may be small wave form that affects the judgment of object locations. Therefore, it should be removed or merged according to wave width and its distance to other wave forms before the moving object is finally outlined (as show in Fig. 5(b)).

3. The Approximate Skin Color Detection and Face Preliminary Screening

After the moving object is detected through the color curve proportion difference algorithm, it is required to detect human faces within the scope of moving objects. In human face detection, extraction of skin color is the most important stage. The characteristics of human skin color have been widely used and proven to be effective in many applications, such as human detection, and human tracking. This skin color detection algorithm is adopted because skin color is important information and its color is different from most background objects. Also, human faces are independent on detailed characteristics of faces. Moreover, it is stable and robots to the changes such as rotation. In this paper, the distribution models of skin color in the color spaces YUV and YIQ are established to get the conditions based on which skin color is determined. Then, according to the conditions, the algorithm seeks out regions that may contain human faces from the image, which is suitable for various illuminations.

Here, analyses are carried out of the skin color distribution characteristics of human faces in the color spaces YUV and YIQ. The component $Y$ represents the brightness information of color. $U$ and $V$ components of the color difference signal are orthogonal and called chrominance signals. Chrominance signals $U$ and $V$ form a two-dimensional vector in a two-dimensional space. The phase angle $\theta$ stands for tone and the formulation is shown as follows:

$$\theta = \tan^{-1}(V / U)$$  \hspace{1cm} (5)

The skin colors of faces of Asians, Europeans and Americans range from red to yellow. According to previous analysis results of the sample values of human face images, it can be considered that the
tonal range of skin colors of human faces is between 105 and 150 [7].

Taking the phase angle as the characteristic, face segmentation is conducted by filtering out those backgrounds which are greatly different from the skin colors of human faces in tone. Also, the saturation information of color can be used in the YIQ space to enhance the segmentation effect. The skin colors of human faces contain much yellow component. Component I represents tones from orange to blue-green. The smaller the value of I is, the more yellow is contained and the less blue-green is contained. Experiments have found that values of skin color of human faces in the YIQ space varies in a specific range which can be determined to lie between 10 and 100 [7].

The range obtained above is just a rough one. We need more accurate thresholds to select skin color regions. Therefore, what will be done next is to choose more accurate skin color regions by establishing skin color models.

Sampling of skin color is conducted as follows:

a. Collect the frontal view human face photos of different genders and ages.

b. Cut the photos manually to get face images.

c. Data analysis is carried out of the great number of manually cut face images, including getting the maximum value $T_{\text{max}}$ and the minimum value $T_{\text{min}}$, the mean square error $E_T$, the mean value $E$ and the best threshold of the skin color of each image $T_1$, $T_2$, $T_0^1$, and $T_0^2$.

Under normal illumination conditions, let $T_1=30$, $T_2=95$, $T_0^1=100$ and $T_0^2=135$, where $T_1$ refers to the lower threshold of I; $T_2$ refers to the upper threshold of I; $T_0^1$ refers to the lower threshold of $\theta$ and $T_0^2$ refers to the upper threshold of $\theta$. Statistics have indicated that due to the influence of illumination intensity, proper adjustments should be made of the parameters based on the summary of the data about the established skin color model.

First set the value of $T_1$, and then set the value of $T_2$ based on $T_1$.

(A) When $T_{\text{max}}>95$, $T_2=95$;

(B) When $T_{\text{max}}<95$, $T_2=T_{\text{max}}$;

a. When $T_1>65$, $T_{\text{max}} \geq 85$ and $ET<20$, $T_1=20$;

b. When $30 \leq T_1 \leq 65$;

1. When $E \leq 5$, $T_1=15$;

2. When $E>5$, $T_1=25$;

c. When $T_2 \leq 30$, $T_1=T_{\text{min}}$;

After proper adjustments of parameters, $T_1$, $T_2$, $T_0^1$, and $T_0^2$ of each image can be obtained. With these four thresholds, approximate skin color regions and non-skin color regions can be distinguished easily in Fig. 6.

4. Structure-Based Frontal Face Detection

After extraction of approximate skin color, we judge whether the window images belong to human faces through the structure-based frontal view human face detection method [8, 9].

![Fig. 6. Approximate skin color region.](image)

1) Some principal organs such as eyes, nose, mouth, etc. are of low gray scales, including the extreme points on curved surfaces. All these are key feature points describing human faces;

2) Except these principal organs, the gray scale of other parts of human faces changes gradually. Fig. 7 is the gray scale distribution map of frontal view human face images.

![Fig. 7. The structural characteristics of the front face.](image)

As human face characteristic distribution has this characteristic, we combine several methods to conduct second identification of human faces. First, preprocessing is conducted of approximate color skin regions, including normalization and background filling so as to facilitate the further processing of the images. Then we make a mosaic image. Then, in order to locate the five sense organs, horizontal line detection from the mosaic image is conducted by using edge detection operators. Finally, based on the approximate skin color detection module, the candidate human faces with different sizes are segmented, in which skin color accounts for a large portion and the background accounts for a small portion. So it is essential to conduct preprocessing, including normalization and background filling. Background filling is about filling some parts of the background based on the average color of the proposed parts so that its color is similar to skin color and that the facial characteristics are further highlighted for easy search.

Mosaic of the preprocessed images is made to reduce the resolution of images. The image shows a common feature of two eyebrows, two eyes, the noise and the mouth on human faces. That is, as Fig. 7 shows when the observed human faces are not very
oblique, they basically present horizontal distribution, which is even more obvious in mosaic images. Therefore, mosaic images before horizontal line detection not only can highlight the characteristic, but also can further reduce the computations of this algorithm. Next, horizontal line detection of images can be conducted using edge detection operators so as to further locate the five sense organs. Sobel operator for edge detection can be used to extract grooved horizontal lines. This method is simple with good effects. Although the extracted edges are thick, they have no significant influence on the algorithm. The horizontal lines of human face organs are relatively long, so their edge detection operators can be used.

After the stripes detection, color image turns into a binary image composed of a number of stripes, where the stripes are the facial features and some interfering elements of the edge line as Fig. 8 shows. According to these different thresholds binaries, the mutual position of the stripes and lines and their relationships, we can further locate the position of the facial features. In the facial features, the most significant is the eyes, and it is also the largest organ with the whole face difference. Therefore, the first detection of binocular binaries by higher draws decreases the number of stripes and makes it easy to reduce the computational complexity.

After stripes detection, color image turns into stripes and forms a binary image, where the stripes, the edge line of the facial features, interference elements, and then different threshold binaries as shown in Fig. 9(a) and Fig. 9(b). \( L1 \) is the result after the high threshold value binary as show in Fig. 9(a) shows, while \( L2 \) is the result after the low threshold value binary as Fig. 9(b) shows.

In accordance with relationships of the mutual position of these stripes and lines, we can further locate the position of the facial features. In the facial features of the human face, the most significant is the eyes, and it is also the largest organ with the whole face difference. Therefore, first binocular testing, used by the detection of the eyes in the figure is \( L1 \), which is drawn by higher binarization threshold. With fewer stripes, it is easy to reduce the computational complexity.

This detection algorithm first searches the coordinates corresponding to the location of eyes based on the length-width ratio of the eye and the relative location of two eyes. Then the point with the minimum pixel is based on to find the coordinates near the pupils. Next, remove the eyebrow pair above the eye pair to make preparation for better search of other characteristics. Finally, the threshold low \( L2 \) image is used to search for the nose and mouth. The two eyes, the nose and the mouth are combined into a group to identify human faces for the candidate five sense organ pair.

According to the above procedures, the location information of the organs on human faces is obtained. Then second identification is applied to determine whether the recorded is the human face.

As we can see in Fig. 10, A, B and C are the extreme points of the mouth region respectively. The matrix region determined by A and C is called \( R1 \); the region determined by B and C is called \( R2 \). Then we make binary images of human faces. \( K1 \) and \( K2 \) are defined as the proportion of white points in \( R1 \) and \( R2 \), respectively. The location of B and C reflects the first feature of human face structure. That is, a certain geometrical relationship exists among main organs. \( K1 \) and \( K2 \) describe the second feature of human faces and respectively reflect the smoothness of regions \( R1 \) and \( R2 \). Therefore, we adopt rules to describe the structural characteristics of human faces: Rule 1:

1. They constitute an acute triangle;
2. A and B basically lie on the same horizontal line;
3. Horizontally, C lies between A and B;
4. Vertically, C lies below A and B;

Rule 2:

1. \( K1 \) and \( K2 \) reflect the smoothness of the regions \( R1 \) and \( R2 \); they must comply with the following equation:

\[
K1 \geq 0.6 \tag{6}
\]

\[
K2 \geq 0.6 \tag{7}
\]

\[
|K2 - K1| \leq 0.2 \tag{1}
\]
The structure based frontal view human face detection algorithm inputs a grayscale image. If the input image belongs to human face model, the structural characteristics of human faces are met; otherwise, the image is not a human face, which is described in detail as follows:

1) Conduct regional segmentation of the input image to obtain feature points A, B and C;
2) If features A, B and C do not satisfy rule 1, the input image is not a human face;
3) If $K_1$ and $K_2$ do not satisfy rule 2, the input image is not a human face;

The key to the algorithm pretreatment step is locating feature points A, B, and point C, and the methods used are: first making regional segmentation of the gray-scale smoothing image and then been able to roughly reflect the gray distribution structure in split image; then obtaining the minimum point of the gray scale value in each region, i.e. one feature point as a candidate feature point; finally, selecting three of the candidate with the smallest gray value eigenvalue as feature points, to ensure the human face. Fig. 11 is some face secondary identification results.

5. Experimental Results

To evaluate the performance of the proposed method, the total number of test samples is 250 video frames, which includes a complicated background with single or multiplayer image. All the experiments were run on a dual core (TM) 2.4 GHz machine with 2 GB of main memory and a Windows XP operating system.

With the complex background and slowly-changed illumination, we have tested 355 moving objects in 250 video frames. We use our algorithm and face detection method in reference [4] to do comparative experiments.

Because in reference [4], it directly subtracted background frame and prospects frame to get moving object and also use simple skin color model to detect human face. To contract with it, we can know the algorithm we propose can solve some problem that similar algorithm didn’t solve. The comparison results are as followings.

First, in the comparison test of moving object detection algorithm, our algorithm is correctly positioned 303 moving objects, the accuracy was 93.4 %. The reference [4] correctly positioned 284 moving objects. The accuracy was 87.7 %. Our algorithm accuracy has improved greatly. Table 1 shows the result.

The comparison test results show that our algorithm has a high sensitivity for indoor moving objects. Although the time interval between background frame and prospects frame is short, the light still slowly changes. If directly subtracted as reference [4], these minor differences will also cause a lot of noises. Thus affecting the accuracy of
positioning moving objects. This problem can be solved by using the method of this article, so as to achieve accurate positioning.

Secondly, we also use the above test samples to evaluate the whole dynamic face detection system. In this paper, face detection accuracy was 89.8 %, while the reference [4] was 83.8 %. Specific comparison test results are shown in Table 2.

The test results show that the algorithm can achieve a better detection of single face in complex background. The inaccurate position to the moving objects cause some of the human faces missing. Therefore it reduces the accuracy of the entire face detection system. So the moving object detection accuracy has a certain effect on face detection accuracy. The accuracy improvement on moving object detection helps to improve accuracy of face detection.

### Table 1. The comparison test result of moving object detection.

<table>
<thead>
<tr>
<th></th>
<th>Our algorithm</th>
<th>Reference [4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving objects</td>
<td>335</td>
<td>335</td>
</tr>
<tr>
<td>The number of undetected</td>
<td>22</td>
<td>41</td>
</tr>
<tr>
<td>Accurate positioning</td>
<td>313</td>
<td>294</td>
</tr>
<tr>
<td>Wrong positioning</td>
<td>33</td>
<td>51</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>93.4</td>
<td>87.7</td>
</tr>
</tbody>
</table>

### Table 2. The comparison test result of the human face detection.

<table>
<thead>
<tr>
<th></th>
<th>Our algorithm</th>
<th>Reference [4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face numbers</td>
<td>335</td>
<td>335</td>
</tr>
<tr>
<td>The number of undetected</td>
<td>32</td>
<td>54</td>
</tr>
<tr>
<td>Accurate positioning</td>
<td>301</td>
<td>281</td>
</tr>
<tr>
<td>Wrong positioning</td>
<td>40</td>
<td>61</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>89.8</td>
<td>83.8</td>
</tr>
</tbody>
</table>

Also the established skin color model and the structure-based frontal view human face detection method help us to detect human face more accuracy.

### 6. Conclusion

Dynamic face detection used to monitor and record the abnormal entry of personnel is a critical step. Considering the difficulties of illumination changes, skin color detection under different lighting conditions and objects with approximate human face structure, this paper proposes a new dynamic face detection algorithm, which is robust to the slowly varying illumination effect. We build the skin color model by statistics based on a large amount of color data to get the approximate skin color accurately, using the structure-based frontal view human face detection method to exclude the wrong face and get the right one.

### References