Face Location in Video Sequence with Complex Background

Zhangli Lan[1,2] Shuang Liang[1] and Jianqiu Cao[1] XiaoDong Liu[3]

[1]Chongqing Jiaotong University, Chongqing, China

[2]Chongqing University, Chongqing, China

[3] Faculty of Urban Construction & Environmental Engineering, Chongqing University, 400045, China

ABSTRACT: The application of face location and recognition in smart building is very important. In order to improve the accuracy of face location in video sequence, the paper presents a new method for rapidly and accurately locating face in video sequence through analysis of face features in video sequence. Background removing, skin region detecting, gradient calculating and feature component fitting technology are adopted in this method. The experiments demonstrate that this method can meet the requirements of actual application. In addition, face-based access control system has been developed using this method.

Keywords: Smart Building, Access Control, Face Location, Video Sequence, Complex Background

INTRODUCTION

With the development of computer technology and smart buildings, computer technology has been continuously incorporated in the new type of smart buildings. Recent years, mode recognition-based intelligent recognition has been rapidly developing and gradually applied to intelligent access control system and security surveillance of new building types, such as application of finger print recognition, iris recognition and face recognition in access control systems. Because finger print recognition can not achieve good result in recognition of children under 6 years old and iris recognition cause some damage to eyes, the scope of application of face recognition has been expanded and research on face recognition has been deeply developed.

As a basis of face recognition and comparison, face detecting and location is an important research field. Only when face location is detected, face recognition can be well done. At present, there are three methods for detecting face, one method is moldplate-matching method based on entire face, which requires large amount of calculation and very easily affected by variation in brightness [1]. The second method is artificial neural networks method. The method requires many sample training activities and then various image regions are eventually judged. This method is feasible theoretically, however, selection of training samples and network convergence are very difficult. As a result, it is difficult to apply this method widely [2]. The third method is skin detecting method, for which color information in digital image is utilized and detecting speed is improved significantly. However, due to crossing of skin space and other color space, the existing methods are limited to sample background and error rate is very high under complex background.

Focusing on face features in video sequence, a face detecting method using background removing, skin regions detecting, gradient calculating and features component fitting is presented in the paper to rapidly and accurately locate face in video sequence. The relevant experiments show that the method achieved satisfactory results.

The process, methods for various steps and experiment results for the algorithm are explained in the paper.

2. Face detection algorithm

2.1 General structure of the algorithm

The general structure of the algorithm is as shown in Fig. 1.

At first, obtain several consecutive frames of image in video sequence and judge whether image is at standstill state. If so, assume that no person moves in image, and store the current image as the background. If the image is not at standstill state, take one image and perform minus operation between this image and background, in order to eliminate background effect and obtain human body location. Then, search for skin regions among human body locations and find out possible face location as candidate face. Then, in face candidate regions, perform gradient calculation and find and mark the locations with large gradient values. Fit brow, eye, nose and mouth location for these marked points. If the gradient points of brow, eye, nose and mouth having appropriate proportion, confirm that the candidate regions is a face.

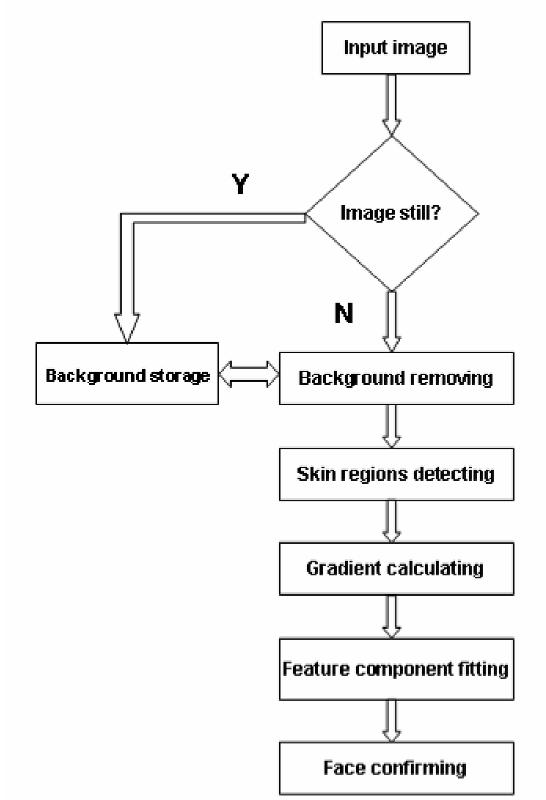

Figure 1: Structure of face detection Algorithm

2.2 Background estimating and removing

Detecting complex background is an important work for face locations. In complex background, some objects have the color similar to skin regions and feature points similar to face. Such objects easily result in incorrect face location. As a result, it is very important to remove interference from background.

Video cameras for access control system of smart building are generally fixed and the images taken by cameras keep stable without change in brightness. Therefore, take stable background image as a reference. Only when image is changed, consider that there are moving objects in image. If only detecting and locating face in moving objects, the interference from background can be completely eliminated.

The detailed steps are as follows:

1. Take several consecutive frames of video image and judge whether any difference exists among these images or not.
2. If no difference exists, it means that there are not any moving objects in current video images, and take one frame as background image. (The reason for taking one standstill image at a time is change in ray during each day, therefore, the background should also change in real-time manner).
3. If any difference exist between many frames, take a frame and minus background image frame. Because the background remains unchanged, the pixel brightness in unchanged background approaches zero after minus operation. And difference exists between moving objects and background and generally far larger than zero. Therefore, it is possible to mark moving objects.

The operation result is as shown in Fig. 2 through 4.

Figure 2: Image of background

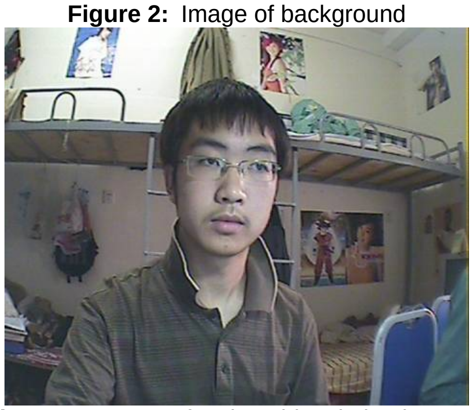

Figure 3: Image of active object in background

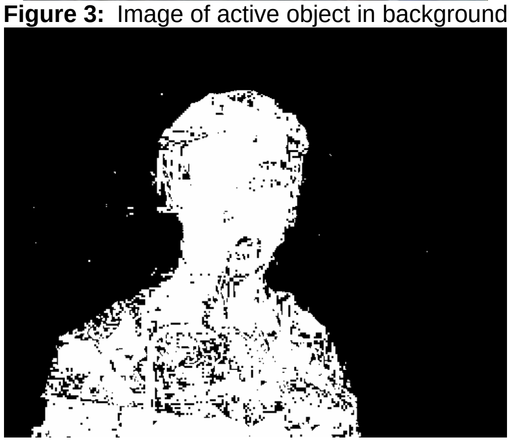

Figure 4: Result of background removing

After background removing operation is performed, it is necessary to only perform subsequent operation for image portion in moving objects. This does not only eliminate interference from background but also reduce the amount of subsequent operations. As a result, this practice can achieve good result.

2.3 Skin regions detecting

After removing background, the moving objects in image have been marked. However, because some colors in moving objects are similar to background, salt-pepper noises exist in marked image. Before detecting skin regions, the marked regions should be opened and closed properly.

After removing background, the range of skin regions detecting is reduced to the regions of moving objects. In this way, interference from background color is avoided, detecting range is reduced and the amount of operations is significantly lowered.

When detecting skin regions, considering that skin difference at different environments is mainly affected by brightness and seldom affected by chroma, ***YC_rC_b*** model is adopted for color matching and detecting to minimize brightness effect upon skin. Generally speaking, the image obtained by computers are of RGB pattern, therefore, Formula 1 is adopted to convert RGB image into ***YC_rC_b*** model.

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ 0.439 & -0.368 & -0.071 \\ -0.148 & -0.291 & 0.439 \end{bmatrix} \times \left(\begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right) \qquad 1$$

Chroma in ***YC_rC_b*** color space converted from RGB contains some brightness elements. For this reason, non-linear chroma conversion is used to solve this problem [3]. This formula is expressed as follows:

$$C_i(Y) = \begin{cases} \left(C_i(Y) - \bar{C}_i(Y)\right) \cdot \dfrac{W_{C_i}}{W_{C_i}(Y)} + \bar{C}_i(K_h) & if \;\; Y < K_l \;\; or \;\; K_h < Y \\ C_i(Y) & if \;\; Y \in |K_l, K_h| \end{cases} \qquad 2$$

$$W_{C_i}(Y) = \begin{cases} WLc_i + \dfrac{(Y - Y_{\min}) \cdot (W_{c_i} - WLc_i)}{K_i - Y_{\min}} & if \;\; Y < K_l \\ WHc_i + \dfrac{(Y_{\max} - Y) \cdot (W_{c_i} - WH_{c_i})}{Y_{\max} - K_h} & if \;\; K_h < Y \end{cases} \qquad 3$$

$$\bar{C}_b(Y) = \begin{cases} 108 + \dfrac{(K_l - Y) \cdot (118 - 108)}{K_l - Y_{\min}} & if \quad Y < K_l \\ 108 + \dfrac{(Y - K_h) \cdot (118 - 108)}{Y_{\max} - K_h} & if \quad K_h < Y \end{cases} \qquad 4$$

$$\bar{C}_r(Y) = \begin{cases} 154 - \dfrac{(K_l - Y) \cdot (154 - 144)}{K_l - Y_{\min}} & if \quad Y < K_l \\ 154 + \dfrac{(Y - K_h) \cdot (154 - 132)}{Y_{\max} - K_h} & if \quad K_h < Y \end{cases} \qquad 5$$

Where ***C_i*** in (3) and (4) is either ***C_b*** or ***C_r***, $W_{c_b} = 46.97$, $WLc_b = 23$ $Y_{\max}$, $WHc_b = 14$, $W_{c_r} = 38.76$, $WLc_r = 20$, $WHc_r = 10$, $K_1 = 125$, $K_h = 188$.

These parameter values are estimated from training samples of skin patches from a subset of the HHI images.

The 2D projection of the transformed ***$C_b'C_r'$*** is in an ellipse [3]. The elliptical model for the skin tones in the transformed ***$C_b'C_r'$*** space is described in (6) and (7)

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1 \qquad 6$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b' - c_x \\ C_r' - c_y \end{bmatrix} \qquad 7$$

Where $c_x = 109.38,$ $c_y 152.02,$ $\theta = 2.53,$ $ec_x = 1.60,$ $ec_y = 2.41,$ $a = 25.39,$ $b = 14.03$ are computed from the skin cluster in the ***$C_b'C_r'$*** space.

The application shows that, when detecting skin regions by using the above elliptical model, some pink non-face regions will be contained in elliptical for skin zone, some relatively-black skin (for an example, at backlighting condition, or deep brownish black skin) is beyond skin ellipse and thus detected improperly. Compensation and removal will be performed respectively when detecting skins.

Fig. 5 and 6 are skin regions found according to the above algorithm. In these figures, the regions, which length-width ratio basically matches the one of faces, are marked by a white rectangle.

Figure 5: Result of skin region detection- Ⅰ

Figure 6: Result of skin region detection-Ⅱ

In Fig. 5 and 6, some color regions meet skin requirements. However, because of existence in the background, these regions will not be calculated and detected after background removing. In addition, some skins are discarded and not marked with white rectangle due to their improper length-width ratio.

2.4 Gradient calculating

It is shown from Fig. 5 and 6 that the range of skin detecting does not only include faces but also other regions matching with skin tones, such as arms shown in Fig. 5 and pattern on clothes shown in Fig. 6. Comparison of face regions with other regions shows that face regions meet the skin requirements, there are brow, eye, nose and mouth at appropriate position on face and these features are unavailable at other regions.

To distinguish between face regions and non-face regions in skin regions, the availability of features points of brow, eye, nose and mouth are used for judgment. It is known from observing these feature points that distinct brightness difference exist between point location and skin color and this is presented by vertical gradients. Therefore, calculate the gradient for skin regions.

Assuming that the position of pixel on image is as follows:

f(x,y)	●	●	***f(x+1,y)***
f(x,y+1)	●	●	***f(x+1,y+1)***

In consideration of the shape characteristics of brow, eye, nose and mouth, the image gradient along vertical direction is biggest and the vertical gradient operator is applied:

$$G[f(x,y)] = \left| f(x,y) - f(x,y+1) \right| \qquad 8$$

The typical face and face gradient are as shown in Fig. 7

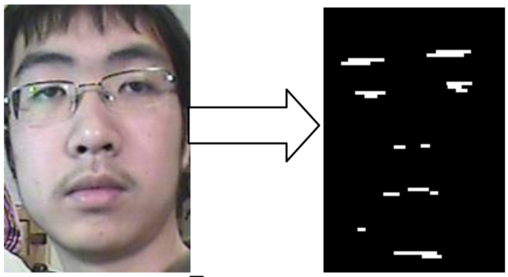

Figure 7: Gradient in a face

Set appropriate gradient threshold according to gradient characteristics of faces. Gradient should exist at brow, eye, nose and mouth and not exist at cheek. A sample face template can be created as shown in Fig. 8.

Face position can be determined on face candidate regions according to the features of face template. The calculation result for Fig. 5 using gradient template of feature points is as shown in Fig. 9.

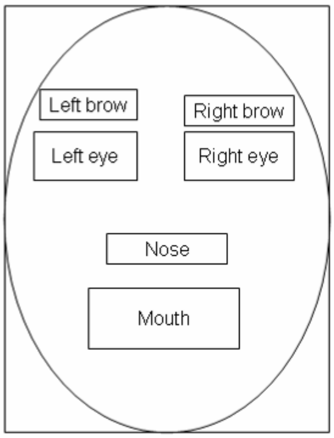

Figure 8: Face template

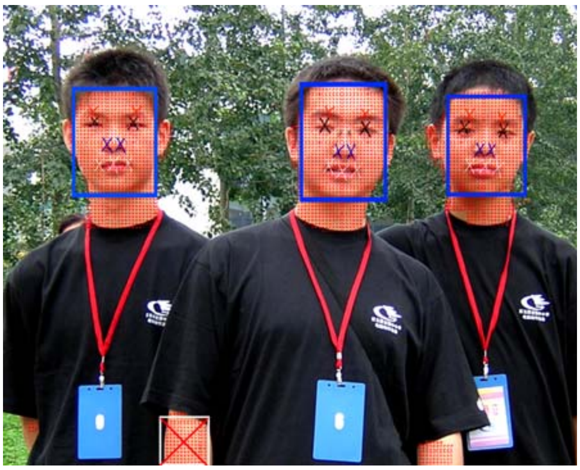

Figure 9: Face confirm with Gradient template

3. CONCLUSION

In the paper, a face location method in video sequence with complex background is researched and described. The experiments shows that, by using the method and Pentium Ⅳ2.8G, it takes about 60ms to simultaneously locate three faces under complex background, the accuracy rate of detecting faces, which tilts or displace by no more than 15 degrees, is more than 97 percent. Therefore, the method can meet the requirement of face recognition-based access control system for smart building. At present, we have developed digital smart multi-function access control and alarm system using the proposed method.

ACKNOWLEDGEMENT

This work is supported by New Century Excellent Talent Funds of Educational Ministry of China (NCET-05-0759), Doctorate Foundation of Educational Ministry of China (20050611001) and Natural Science Foundation Project of CQ CSTC.

REFERENCES

1. Leung T.K,Burl M.C,Perona P, Finding faces in cluttered scenes using random labeled graphic matching, Fifth Intl. Conf. on Comp. Vision, Cambridge .M.A. June, 1995
2. Li Shijin et al. Robust face detection in complex background, Mini-micro Systems, Vol.21,No.7,July 2000.
3. Rein-Lien Hsu. Mohamed Abdel-Mottaleb and Anil K. Jain. Face detection in color image. www.image2003.com