Journal of Korea Multimedia Society Vol. 14, No. 8, August 2011(pp. 992-999)

Face and Hand Activity Detection Based on Haar Wavelet and Background Updating Algorithm

Shang YiTing†, Lee Eung Joo††

ABSTRACT

This paper proposed a human body posture recognition program based on haar-like feature and hand activity detection. Its distinguishing features are the combination of face detection and motion detection. Firstly, the program uses the haar-like feature face detection to receive the location of human face. The haar-like feature is provided with the advantages of speed. It means the less amount of calculation the haar-like feature can exclude a large number of interference, and it can discriminate human face more accurately, and achieve the face position. Then the program uses the frame subtraction to achieve the position of human body motion. This method is provided with good performance of the motion detection. Afterwards, the program recognises the human body motion by calculating the relationship of the face position with the position of human body motion contour. By the test, we know that the recognition rate of this algorithm is more than 92%.The results show that, this algorithm can achieve the result quickly, and guarantee the exactitude of the result.

Key words: Activity Detection, Wavelet, Background Updating

1. INTRODUCTION

Human Detection is an important orientation in the fields of computer vision and pattern recognition. It has an important role in graphic image processing, intelligent control, video coding and other fields.

Using computer vision systems to perform motion recognition is a complex and challenging task. How to capture information on the movement of the body? How to determine the movements for gesture recognition? They are very difficult. This paper builds a body gesture detection system. It contains three main parts: 1) Moving target detection to extract moving regions.; (2) Obtaining the face location using Haar-like features (3) Estimating body posture using the location of face and human body posture.

This paper studies a method based on background subtraction, presenting a moving target detection method based on adaptive background updating. The analysis method is based on a currently used method. This method indicates the background image by improving the mix of a multi-Gaussian model, and background subtraction to detect the moving target region. This paper uses the AdaBoost algorithm using face images of Haar-like features of statistical learning, to generate a strong face detection classifier. The detection uses the AdaBoost detector for color images that may exist in the region in YCrCb color space using the skin color model for face location for precise positioning. Results show the method proposed to detect a moving target can detect the moving regions quickly. It largely suppresses background noise and changes in the region; Face

※ Corresponding Author : Lee Eung Joo, Address : Dept. of Information Communications Engineering, Tongmyong University, 535 Youngdang-Dong, Nam-Gu, Busan 608-711, Korea, TEL : +82-51-629-3700, FAX : +82-51-629-3729, E-mail : ejlee@tu.ac.kr
Receipt date : Mar. 23, 2011, Revision date : May 16, 2011
Approval date : June 23, 2011
† Dept. of Information Communication Engineering, Tongmyong University
(E-mail: yitingshang@nate.com)
†† Dept. of Information Communication Engineering, Tongmyong University

Detection can quickly and accurately detect the face location based on Haar-like Features.

2. AN APPROACH OF FACE DETECTION IN COLOR IMAGES BASED ON HAAR WAVELET

This article uses an YCbCr color space model to perform image segmentation to obtain the color regions. It uses the Adaboost learning algorithm to obtain the face detector based on the Haar wavelet transform, and perform detection of the skin color region; on the face region, using the eye color and mouth features extract the eye and the mouth candidate regions, respectively; finally, the geometric structure of the relationship between facial organs using the template matching method identifies the location of eyes and mouth.

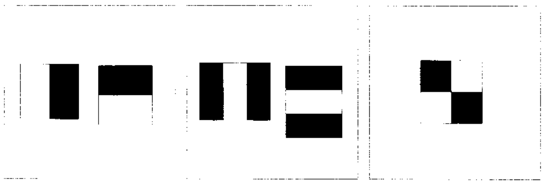
2.1 Skin color segmentation

The RGB model is more consistent than description of facial skin color. Brightness and chroma have little correlation in the YCbCr color space; the RGB model is more consistent than the description of facial skin color. RGB conversion is relatively simple. The image from RGB space to YCbCr space mapping Cb, and Cr components of a two-dimensional plane, based on a skin color model, can determine the candidate region through a number of morphological closings and openings. Noise is removed and small area is filled with large holes to obtain regions and the corresponding cover graph.

2.2 Face detection based on haar feature

Papageorgiou [1] used the Haar wavelet transform to extract face features from local Haar features. The algorithm used three types of local Haar features.

The characteristic value of each class is all the pixels within the white rectangle and the gray value, minus the black rectangle in the region and all the pixel gray values (Figure 1).



(a) (b) (c)

Fig. 1. (a) Boundary (b) Leptonema (c) Diagonal.

Viola [3-5] proposed a fast algorithm in which an integral image can easily be extracted from the characteristics of the local Haar features. The image coordinates of points are defined in the point image from the point at the top left. All the pixel gray values are expressed as line on the pixel gray value. Using formula (1), (2), a traversal of the original image can calculate the integral image of all points. Using the integral image, any of the original image pixel gray values within a rectangle computation is required as a constant, which can quickly calculate the characteristics of each feature value (Figure 2 and formula (3)).

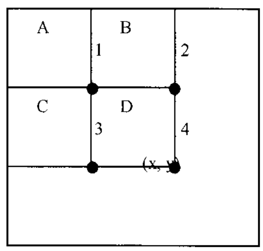


Fig. 2. Schematic drawing.

$$ii(x,y) = \sum_{x'\le x,\, y'\le y} i(x',y') \tag{1}$$

$$s(x,y) = s(x,y-1) + i(x,y)$$
$$ii(x,y) = ii(x-1,y) + s(x,y) \tag{2}$$

$$valueD = ii4 - ii1 - ii2 - ii3 \tag{3}$$

We obtain the face images to make a face that contains 500, 1500 non-face images in a training sample set, each sample size is 24 × 24 pixels. For each feature, calculate the corresponding sample eigenvalue; it will have the minimum classification error rate of the characteristic value as a threshold, to obtain a weak classifier:

$$h_j(x)=\begin{cases}1 & p_j f_j(x)< p_j\theta_j \\ 0 & otherwise\end{cases} \quad (4)$$

Expressing a 24 × 24 image of the sample, the feature of the characteristic value, is a dual factor. In contrast to the method used by Viola, we obtain training samples directly. We assign the face sample a large weight. In addition, we assign the non-face sample a small weight. Thus, at the stage of a weak classifier in the training, we focus on the face samples. We obtain the weak classifier with better classification results, reducing the use of the number of strong classifiers in the weak classifier, thus improving the detection rate.

2.3 Human eye and mouth detection

We source 100 people from the photo gallery of the Heinrich-Hertz-Institute (HHI) to establish the color model of the human eye. Then, we manually remove the image in the eyes, because the human eye whites have the same color. When selecting the eye area, we retain only part of the cornea, to get rid of highlights produced by the corneal reflex. Finally, we obtained a 22,000 pixel sample set. We perform RGB → YCbCr color space transformation for the sample set. As can be seen, the eye color in the YCbCr color space distribution is more concentrated. Most Y component samples are concentrated in (0,120). In addition, the Cb Cr component is generally a higher value component. From the above analysis, in the color space, eyes and skin Cb, Cr components are quite different, from which the location of the human eye can better determined.

We can use equation (formula 5) to strengthen the eye area, because the human eye has a strong component:

$$EyeImageC=\left(Cb^2+(255-Cr)^2+(Cb-Cr)^2\right)/3$$
$$Cb, Cr \in [0,255] \quad (5)$$

Compared with other organs, the color of the mouth has a higher-intensity Cr component and low intensity Cb component. Thus, we can use non-linear processing (formula 6) to enlarge Cr and Cb components of the difference between intensities to highlight the mouth region area.

$$MouthImage = Cr^2\left(Cr^2-\eta Cr/Cb\right) \quad (6)$$

$$\eta = 0.95*\frac{\sum_{(x,y)\in f} Cr(x,y)^2}{\sum_{(x,y)\in f} Cr(x,y)/Cb(x,y)} \quad (7)$$

Measurements found the human eye and mouth area of the face were about 3% of the entire area. Therefore, we extract the human eye and mouth to reserve 3% of the maximum gray value of pixels, using the morphology region closing and opening to remove small areas of noise. Then, according to the geometric characteristics of each connected region, such as the size of the region aspect ratio position, to remove the areas that cannot be the human eye or mouth, to obtain EyeImage and MouthImage. Two vertices must be from the EyeImage, another from the MouthImage.

2.4 Human face acknowledgement

We conduct an operation with EyeImage and MouthImage to obtain the feature image, and to determine the center of each connected region. Triangulating a structure from these centers, form the bases of the geometry of the face in the eyes and mouth, to generate a triangular template (Figure 3). We calculate the similarity final score of the triangle with the template (formula 8)~(formula 10), based of the size similarity that we ultimately judge.

$$f_1(\theta_1)=1-\frac{(\theta_1-0.8237)^2}{0.6785} \quad 0\le\theta_1<\frac{\pi}{2} \quad (8)$$

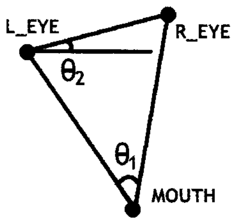


Fig. 3. FaceTriangle model

$$f_2(\theta_2)=\begin{cases}\exp\left(\dfrac{-\theta_2^{\,2}}{0.1921}\right) & 0\le\theta_2<\dfrac{\pi}{2}\\ 0 & else\end{cases} \tag{9}$$

$$finalScore = f_1(\theta_1)f_2(\theta_2) \tag{10}$$

We tested people of different color images (Figure 4). The face detector can accurately detect the face region image combined with the color information used to train the Adaboost learning algorithm. For each rectangular region, this method effectively excludes most of the non-face region. We find the candidate eye and mouth positions in the face in the region. Finally, we accurately determine the position of the face through the similarity calculation and the template.

(a) (b)

Fig. 4. (a) Face (b) Face picture with EyeImage and MouthImage

3. HUMAN HAND MOTION DETECTION BASED ON DIFFERENCE BACKGROUND IMAGE

3.1 Improved real-time background updating algorithm

The background image difference algorithm is based on the static background, but in the real situation, the background is not completely static; it is affected by changes in lighting and weather. Changes in the background images of humans participating in sports have caused great difficulty. Therefore, the background image must be updated in real time.

The study found that under ideal circumstances, adjacent to the background image in the same region, the gray value difference is zero. However, due to the presence of noise and the background slowly changing, the gray value difference is not entirely zero. We analysis changes in the amount of time for the gray value of a pixel to change with no background movement in a static scene. We find the variation of pixels in two adjacent frames in the gray value region. This follows the zero-mean normal distribution $N(0, \sigma^2)$.

$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp(-\frac{x^2}{2\sigma^2}).-\infty<x<+\infty \tag{11}$$

The probability density function is a Normal distribution of mean zero, and average variance of two.

According to the normal distribution "3σ Principle", we can know the probability of a variable value is 99.7% within this range. A pixel in the static scene with no background movement, has an absolute value between two adjacent frames in the gray value of the difference greater than 3σ probability of 0.3%. In this paper, using the normal distribution characteristics, we improve the traditional background updating algorithm. We propose a real-time background update algorithm. The algorithm is as follows:

(1) When no movement human body appears, we study the background to obtain the gray value of each pixel in the sequence of changes in the sample of a number of consecutive image frames, estimating the variance of the image sequence corresponding to pixel gray value. The variance of the K pixel gray value in image sequences is σ_k.

(2) Calculate the value of each pixel in the difference of the current image and background image detection:
D_k is the difference of the pixel gray value k, I_k is the gray value of pixel k in the current test image I, B_k is the gray value of pixel k in the background image B.
(3) When $D_k \leq \sigma_k$, that is, caused slowly changing lighting or weather, at this point, update B_k with I_k.
(4) When $D_k \leq 3\sigma_k$, there is movement caused by the human body, at this point, maintain B_k unchanged.
(5) When $\sigma_k < D_k < 3\sigma_k$, there may be movement in the human body, or background changes; at this point, maintain B_k unchanged.
(6) When B_k is not updated for A long time, update B_k with I_k.
(7) When a large number of pixels of gray level difference $> 3\sigma$; that is, the background image due to special reasons, such as background lighting suddenly turned on or off, camera movement, or sudden changes in weather, causes the image to undergo a large change in the active area; at this time, update the original B with I.

The improved real-time updating algorithm to compute the background is fast, and is update effectively. It can effectively eliminate the affects due to lighting or slow changes in the weather, changes due to background and background mutation.

3.2 Background image difference algorithm

The background image difference algorithm is commonly based on gray images of the motion detection algorithm. Its features are precise location, operation speed. It can split the moving object, but the algorithm is sensitive to changes in background image; it needs to update the background. The background image difference algorithm has the following formula:

$$ID_{(x,y)} = | I_{(x,y)} - BN_{(x,y)} | \qquad (12)$$

$ID_{(x,y)}$ is the gray value after the current test image difference Y Department. $I_{(x,y)}$ is the gray value for the current test image in (x, y), $BN_{(x,y)}$ is the gray value updated background image in the (x,y).

3.3 Based on average gray value image segmentation algorithm

A different image is necessary after image segmentation. The segmentation algorithm is the most common parallel direct detection of region segmentation. The image is divided into foreground and background. Threshold segmentation is equivalent to the image binarization. In essence, each pixel determines a threshold Tk; according to the threshold Tk, determine the pixel k as the prospective pixel or background pixels. Choosing the threshold is difficult. The fixed threshold segmentation algorithm is simple, but the segmentation result is unsatisfactory. The histogram threshold segmentation algorithm needs to have clear peaks dividing the image histogram. When it is not bimodal, it cannot be used.

Using principle 3σ of the normal distribution, consider the selection of pixel k, as a threshold in two adjacent frames in the gray value of the difference, three times the value of the changes in the distribution variance. It was found that, in the difference image, the relationship between the human movement gray value pixel image region and the current average gray value detection image is direct. When the average gray value of the forward test image is large, the gray value difference of the regional pixel of human movement mean is large; when the average gray value of the forward test image is small, the gray value difference of the regional pixel of human movement mean is small. Select the threshold, $T_k = 3\sigma_k$, for image segmentation, when the lighting is dim; the current test image mean gray value is small, so the human movement image area will appear as large areas of rupture and leakage. The prospect is for a large number of pixels to be misidentified as background pixels.

Therefore, this paper proposes a threshold algorithm based on the average gray value image. It uses the product of the current average gray value detection image L and the ratio of coefficients after differential threshold image segmentation:

$$T = \mu \times L \tag{13}$$

T is the threshold difference image segmentation, μ

is the scale factor, L is the average gray value of the current test image.

The experimental study found that when μ is in the range (0.15, 0.3), in conditions of low light and strong lighting, the resulting image segmentation is very satisfactory.

4. HUMAN ACTIVITY ESTIMATION

The calculation of the arm location is the focus of this paper. It is based on the location of the face and arm movement position. This article uses a combination of the position and movement of the face method. This can achieve a better recognition effect.

First, calculate the location of the center of the face and the location of the center of arm movement. If no arm movement is detected, only the detected face shows. When arm movement is detected, we use the center of the face for the y-axis and the arm for the x-axis center to be two right angles right at the triangle's edge. Calculate the distance of the hypotenuse of the center of the face and the center of the arm, when the center of the face and the center distance of the arm is greater than the distance of the increase in the radius of the face and arm. This explains that the arm is above the face; it explains that the arm is moving up. When the center of the face and the center distance of the arm is less than the distance of the increase in the radius of the face to arm, this explains the arm is above the face, and the arm is moving down.

5. RESULTS

For achieving the accurate data about the performance, we respectively tested the program in complex background and simple background. There are four kinds of human motions detection image respectively showed in Fig. 6 in simple background.

This article proposed a body posture recognition method based on face recognition and motion detection. First, we use Haar-like features to obtain the location of the face. Then, using the background subtraction method to obtain the position of human body movement, according to the relationship of the position of the face and arm movement, calculate the body posture of the movement. This algorithm can quickly and efficiently detect the human face. It determines the location of the eyes and mouth. The background subtraction

(a) (b) (c)

Fig. 5. (a) Original image (b) Frame difference image (c) Binary image of motion detection.

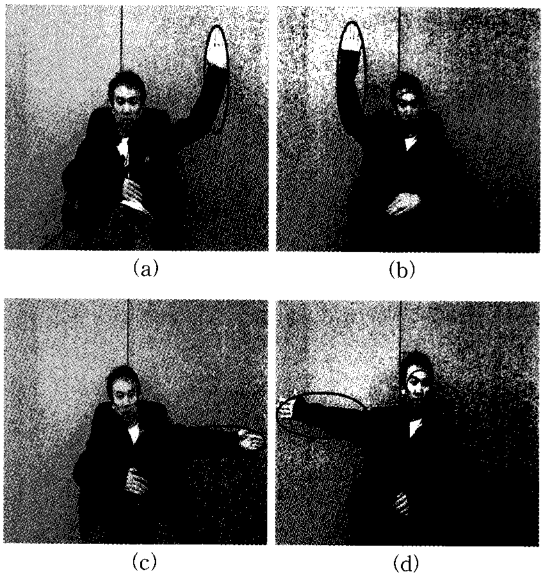


Fig. 6. (a) Left arm raise (b) Right arm raise (c) Left arm lateral raise (d) Right arm lateral raise.

method based on detection of human motion is useful to eliminate effects of noise and shadows. There are four kinds of human motions detection image respectively showed in Fig. 7 in simple background.

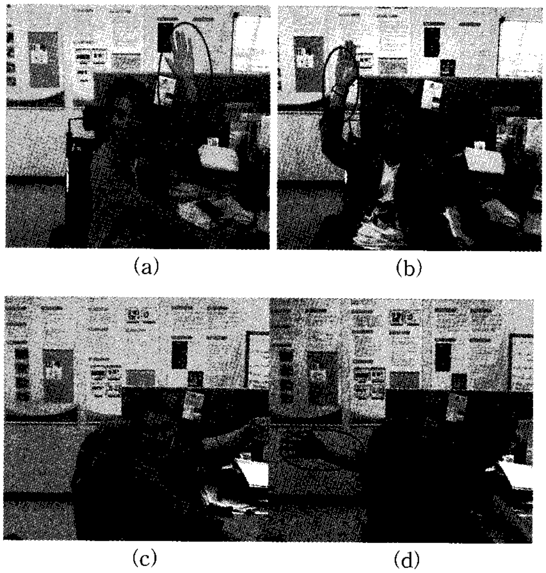


Fig. 7. (a) Left arm raise (b) Right arm raise (c) Left arm lateral raise (d) Right arm lateral raise.

We selected ten test persons, testing them under different background. We found under simple background, human body posture recognition accuracy exceeds 95%. Under complex background, the human face recognition rate can exceed 94% and action detection can exceed 93%. Although face recognition and motion recognition has higher accuracy, human body posture recognition accuracy exceeds 92%.

Table 1 compares the present activity methods based on HMM[6-11] metric learning[10,12,13] and our method. The rate of our method is more then a percentage point above the average rate of other methods in the different background at least. The results are showed in Table 2.

Table 1. Comparison of activity recognition models

	HMM	Metric learning	Our method
Simple background COR	92.1%	91.9%	95%
Complex background COR	91.2%	90.9%	92.5%

Table 2. Human body posture recognition results

	Under simple background		Under complex background	
	Num	Percent	Num	Percent
Left raise COR	196	93%	186	93%
Left raise ERR	4	7%	14	7%
Right raise COR	187	95%	186	93%
Right raise ERR	13	5%	14	7%
Left lateral raise COR	193	92%	184	92%
Left lateral raise ERR	7	8%	16	8%
Right lateral raise COR	184	94%	184	92%
Right lateral raise ERR	16	6%	16	8%
Total COR	760	95%	740	92.5%
Total ERR	40	5%	60	7.5%
Total Test	800	100%	800	100%

REFERENCES

[1] U.P Mosimann, R.M. Muri, D. J. Burn, J. Felblinger,J. T. O'Brien, and I. G. McKeith, "Saccadic Eye Movement Changes in Parkinson's Disease Dementia and Eementia with Lewy Bodies," Brain, Vol.128, No.6, pp. 1267-1276, 2005.

[2] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja, "Detecting Faces in Images: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.1, pp. 34-58, 2002.

[3] Constantine P. Papageorgiou, and Michael Oren, Tomaso Poggio, "A General Framework for Object Detection," Sixth International Conference on Computer Vision, pp. 555-562, 1998.

[4] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," Conference On Computer Vision and Pattern Recognition, Vol.1, No.1, pp. 511-518, 2001.

[5] Paul Viola and Michael J. Jones, "Robust Real-time Object Detection," Cambridge Research Laboratory, Technical Report Series, 2001.

[6] Zhu Y, "Using Moment Invariants and HMM in Facial Expression Recognition," Patern Recognition Leters, pp. 83-91, 2002.

[7] Guo Jing, Rajan D., Chng Eng Siong, "Motion Detection with Adaptive Background and Dynamic Thresholds," Conference on Information, Communications and Signal Processing, 2005 Fifth International, pp. 41-45, 2005.

[8] Guo Jing, Chng Eng Siong, Rajan D., "Foreground Motion Detection by Difference-Based Spatial Temporal Entropy Image," TENCON 2004. 2004 IEEE Region 10 Conference, Vol.1, pp. 379-382, 2004.

[9] Jiangbo Lu, Lafruit G., Catthoor F., "Fast Reliable Multi-Scale Motion Region Detection in Video Processing," Conference on Acoustics, Speech and Signal Processing, Vol.1, pp. 689-692, 2007.

[10] Kalviainen H., "Motion Detection Using the Randomised Hough Transform: Exploiting Gradient Information and Detecting Multiple Moving Objects", IEEE Proceedings Vision, mage and Signal Processing, pp. 361-369, 1996.

[11] Neil Robertson, Ian Reid, "A General Method for Human Activity Recognition in Video", Computer Vision and Image Understanding, Vol.104, pp. 232-248, 2006.

[12] Du Tran, Alexander Sorokin, "Human Activity Recognition with Metric Learning," COMPUTER VISION - ECCV 2008, Vol. 5302, pp. 548-561, 2008.

[13] Bai yu, S.Y. Park, Y.S. Kim, I.G. Jeong, S.Y. OK, and Eung-Joo Lee, "Hand Tracking and Hand Gesture Recognition for Human Computer Interaction." *Journal of Korea Multimedia Society*, Vol.14, No.2, pp. 182-193, 2011.

Shang Yiting

received his B. S. at Dalian Polytechnic University in China (2005-2009). Now he is a M. S student of Tongmyong university. His main research tops are image processing, computer vision, and pattern recognition.

Eung-Joo Lee

received his B. S. , M. S. and Ph. D. in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, and Aug. 1996, respectively. Since 1997 he has been with the Department of Information & Communications Engineering, Tongmyong University, Korea, where he is currently a professor. From 2000 to July 2002, he was a president of DigitalNetBank Inc.. From 2005 to July 2006, he was a visiting professor in the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests includes biometrics, image processing, and computer vision.