

Adaptive Skin Color Detection Based on Human Face under Complex Background

Jing Wan, Ze Yuan & Xuechao Zhang

*College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China
Beijing Technology and Business University, Beijing, China*

ABSTRACT: Traditional skin color detection technique has a weak anti-interference ability against the pixels with color similar to skin under a complex background, and cannot reduce the influence of illumination on the characteristics of skin color. In this study, an adaptive skin color detection method is proposed to tackle the issues. The skin section containing illumination information is extracted by combining the face detection methods proposed by Haar and Adaboost, and using the improved binarization algorithm. Then, combining the best threshold of luminance component (Y) of skin color samples obtained after training in the YCbCr space, the improved histogram backprojection method is adopted to detect the skin color of the whole image. Experiments show that the method is robust under complex background and the influence of illumination. Moreover, the method has a higher accuracy and recall rate than traditional skin color detection methods.

Keywords: Skin detection; Local binarization; Adaptive threshold; Histogram back projection.

1 INTRODUCTION

Skin color detection is an important topic in image processing. It makes contributions to various fields of computer vision such as face detection and recognition, gesture recognition, human-computer interaction and screening for objectionable images based on contents. With the skin color detection technique, we can effectively eliminate the interference from background, reduce the scope of processing, and increase working efficiency and the accuracy of results.

Amidst the studies on skin color detection, Duan et al. [1] proposed a parametric method combining YUV and YIQ color space to fix the color component threshold and to detect the skin color pixels. Peer et al. [2] discovered that the skin color pixels have more superior clustering performance in other color spaces. After transforming the skin color pixels to the YCbCr color space, the components of Cb and Cr will cluster on a 2D plane in the shape similar to ellipse. Accordingly, Peer et al. conducted the fitting of skin color pixels on the Cb-Cr plain to construct an elliptic boundary model, thus achieving the differentiation of skin color and non-skin color. Although the method determining a fixed boundary is convenient and is effective to detect pixels of skin color, the differentiation effect is not satisfactory for the background having similar color to the skin.

To solve the problem of interference caused by complex background, adaptive skin color detection methods [3, 4, 5] have been proposed continually in recent years. In these methods, facial skin color pixels were used as the learning sample, and the Gaussian model or the weighted average Mahalanobis distance was adopted to judge whether other pixels are from

skin or not. However, the methods for extracting skin color pixels were not satisfactory. Literature [4] adopted fixed threshold to extract skin color pixels, while literature [5] applied the Sobel edge detection. When extracting skin pixels with the two approaches, the robustness of shadow processing was not good, and the information was not correct. In the process of adaptive learning, the training of Gaussian model will be time-consuming, and cannot be used for real-time demands. Besides, the extraction of skin color pixels is the training of small sample size. Gaussian model, which is suitable for big sample environment, does not fit this situation.

In addressing the above issues, we tried to improve the adaptive skin color detection methods in this study. First, the improved adaptive binarization method is adopted to extract the skin color pixels of human face. The influence of shadow is fully considered in the improved binarization, so that the extracted information is complete. Then, the histogram backprojection algorithm of Cb-Cr component in the YCbCr color space is combined with the scope of luminance component (Y) to classify the extracted region. In this way, the luminance information is taken into consideration, which contributes to the robustness of illumination and shadow, and decreases the missing rate. Moreover, the skin section in an image can be detected accurately, which is a solution to the issue of interference caused by background similar to skin.

2 ADAPTIVE SKIN COLOR DETECTION METHOD BASED ON HUMAN FACE

The proposed adaptive skin color detection in this study is to update the existing skin color detection

algorithms and the related thresholds using the confirmable information of the image to be detected. The updating result is adopted for the re-detection of the original image to obtain a more accurate outcome. The framework of the proposed algorithm is shown as Figure 1.

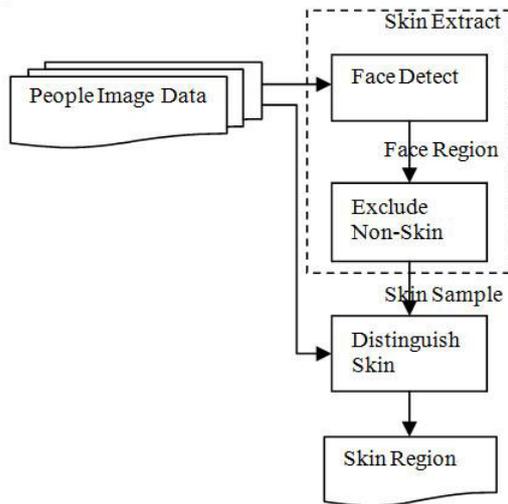


Figure 1. Frame of the adaptive skin-detection

2.1 Extraction of Skin Color Pixel

2.1.1 Face Detection

Accurate extraction of skin pixel samples is the premise of the whole algorithm. To adapt the skin color detection to different environments, the identifiable skin sections are extracted for each image to be detected.

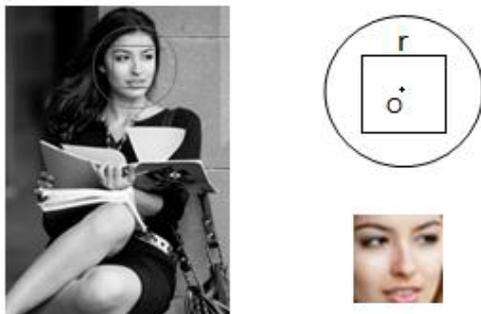


Figure 2. Face detection and segmentation

The objective of face detection is to partition the confirmed skin section and eliminate the interference from the background at the edge and the hair. First, the face detection method of Adaboost and Haar [6] is employed to detect the size and position of the human face. Then, in accordance with the standard face features as shown in Figure 2, a square area covering the

face is drew with the center of the circumcircle of face, O , as the center of this area, and the radius of the circumcircle, r , as the side of length. In this way, the desired objective can be achieved.

2.1.2 Elimination of Non-Skin Section

The non-skin pixels in the area, such as eyebrow, eye, and the shadow of nose, are eliminated. To guarantee the correctness of the extracted skin pixel samples, the face section is divided into three categories: the white of eyes and the teeth; eyebrow, eyeball and the shadow of nose; skin color section. Experience shows that the colors of the above three categories lie in three intervals, which is suitable for comparison and exclusion.

First, the pixels of the white of eyes and the teeth are transformed. The values of three components of white color in RGB space are close and approach to 255, while for the skin color, $R(\text{red}) > G(\text{green}) > B(\text{blue})$ and the values are close. Therefore, the pre-attribute of pixel P can be judged by setting a limiting condition. Where, Value (R), Value (G) and Value (B) denote the values of three components of the pixel in RGB space, respectively.

By annotating the skin sections of 26 face images, 229,736 skin color pixels were obtained. The teeth and the white of eyes in 9 of the images were clear, and 546 non-skin pixels were extracted. By dividing the pixels according to the statistical rules above, it was derived that the two types can be distinguished perfectly with threshold $t=5$.

The next step is to eliminate the eyeball, eyebrow and the shadow of nose. The V component in the HSV space, which denotes the brightness of color and has the range of 0 (black) - 255 (white), is processed. Besides, an improved adaptive thresholding method for binarization is utilized to eliminate the interference of non-skin section. Otsu algorithm (maximization of interclass variance) is the traditional adaptive thresholding method for binarization. It has an outstanding performance on images having a unimodal interclass variance. Nevertheless, if there exists the influence of shadow, the interclass variance will be bimodal or multimodal, and the division performance will be unsatisfactory. As shown in Figure 3, the right part with weak lightness is directly binarized to non-skin color, losing mass effective information.

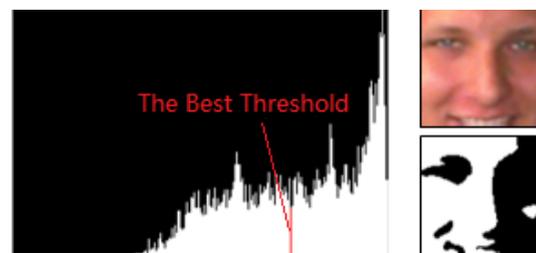


Figure 3. The result of traditional Otsu algorithm

To solve the problem, the local binarization principle is utilized. The face image is divided into 15 rectangular sections by taking the size of an eye as template to form a window, for which the length and height are 1/3 and 1/5 of the length of side of the square area mentioned above, respectively. Then the best threshold, t , is calculated for each of the rectangular section with Otsu algorithm. The process is shown as follows.

Step 1. The image is divided into 15 sections by 1/3 of the width and 1/5 of the height.

$$S_{image} = \sum_{i=1}^{15} S_i \quad (1)$$

Where, S_{image} denotes the input image and, S_i denotes the image after division.

Step 2. The gray-scale histogram of the divided image, S_i , is calculated, and the histogram is subjected to normalization.

$$P_j = \frac{n_j}{M} \quad (2)$$

Where, M is the number of all pixels, n_j is the number of the pixels with gray scale being j , and p_j is the probability that the gray scale of a pixel is j .

Step 3. With the threshold t , the gray scale can be divided into two categories, C_1 and C_2 . The probabilities for the two categories w_0 and w_1 are as follows.

$$w_0 = \sum_{i=0}^t P_i \quad ; \quad w_1 = \sum_{i=t+1}^{M-1} P_i \quad (3)$$

Step 4. The accumulated gray values μ_0 and μ_1 can be calculated according to the probability of each category:

$$\mu_0 = \sum_{i=0}^t (i * P_i) \quad ; \quad \mu_1 = \sum_{i=t+1}^{M-1} (i * P_i) \quad (4)$$

Step 5. The accumulated gray value μ_T for the whole gray scale in the scope of 0-T(255) is calculated as follows:

$$\mu_T = \sum_{i=0}^{T-1} (i * P_i) \quad (5)$$

Step 6. The interclass variance σ is calculated as follows.

$$\sigma = (\mu_T * w_0 - \mu_0)^2 / (w_0 * w_1) \quad (6)$$

The maximal value of interclass variance σ indicates that the difference between the two classes is the greatest. The threshold t corresponding to this situation is the best division threshold.

2.1.3 Effect of Extraction of Skin Color Pixels

Figure 4 is obtained by performing the improved binarization on the face image shown in Figure 3.

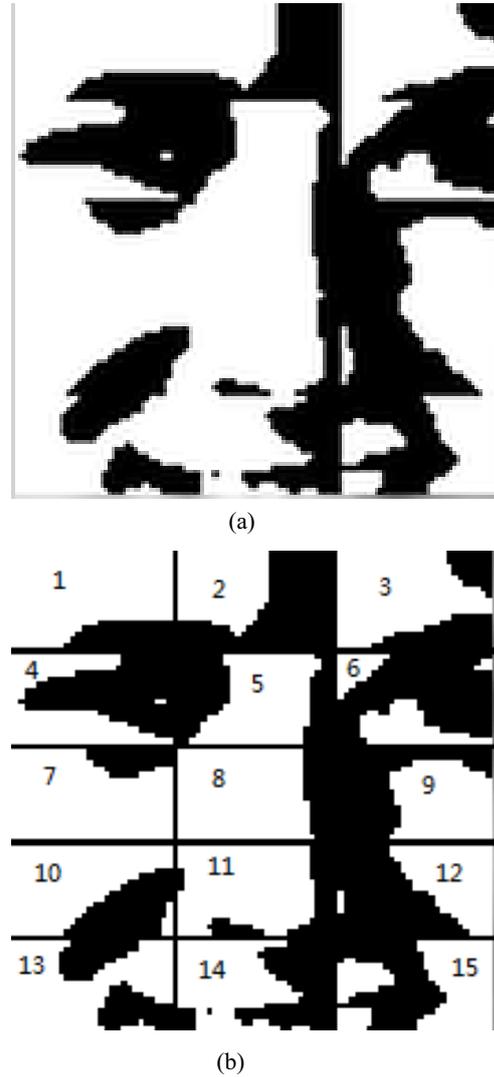


Figure 4. The result of improved Otsu algorithm

The right dark part is clearly improved. However, the sections 1, 2, 7, 10 and 11 in the figure which were previously identified as smooth skin color were divided incorrectly after the binarization, causing the loss of some information. Hence, to guarantee the completeness of information, the lost information should be retrieved by performing geometric operation on the locally binarized result and the globally binarized result, as shown in Figure 5.

2.2 Adaptive Skin Color Judgment

2.2.1 Improved Histogram BackProjection Algorithm

With the blue-difference chroma (Cb) and red-difference chroma (Cr) components in the YCbCr color space as reference, the histogram backprojection



Figure 5. Geometric operate the image

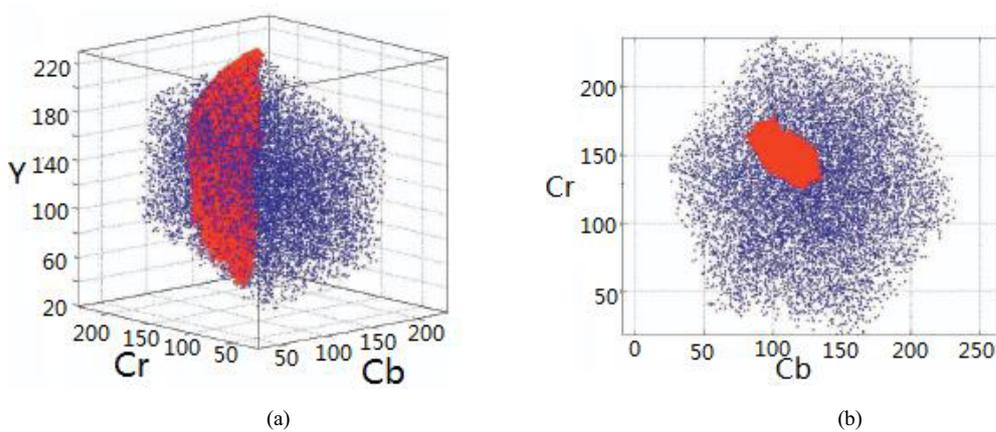


Figure 6. The skin-pixels in the YCbCr space

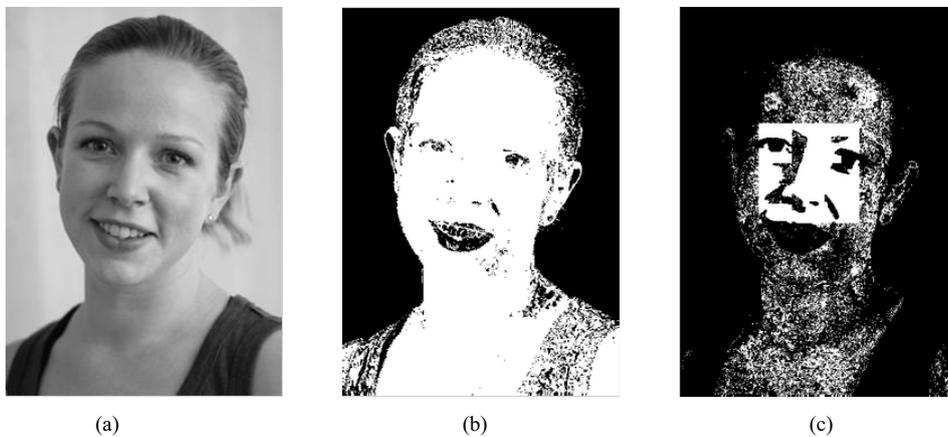


Figure 7. The result of back projection

algorithm is utilized to acquire the skin section and avoid the interference of background. The relevant study result of literature [8] is shown in Figure 6. The clustering performance of skin color pixels is outstanding in the YCbCr color space, especially the projection on Cb-Cr plane. But the Y component has little fluctuation along the Y axis. Therefore, most approaches proposed in previous skin color detection studies were based on the Cb and Cr components. However, the skin pixels distributed in Cb-Cr plain

shown in Figure 6(a) are still influenced by the Y component (the scope of Cb-Cr component is small where the luminance is relatively high and low). If only the Cb and Cr components are considered, the false positive rate will increase, as shown in Figure 6(b). If all color information in Y-Cb-Cr is used as the parameter of back projection, it is equivalent to searching the pixels having the same value as the pixel sample in the image. As a result, a large number of pixels will be undetected, as shown in Figure 7(c).

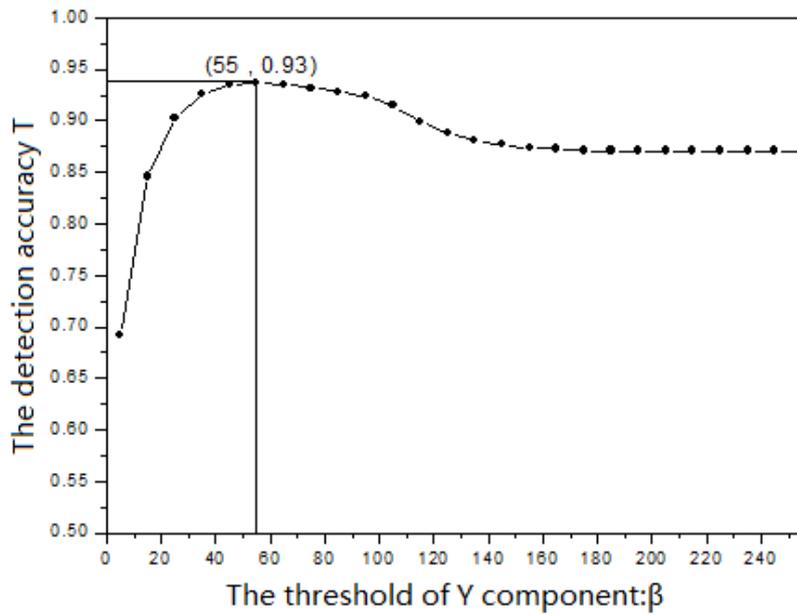


Figure 8. The relationship between the accuracy and threshold

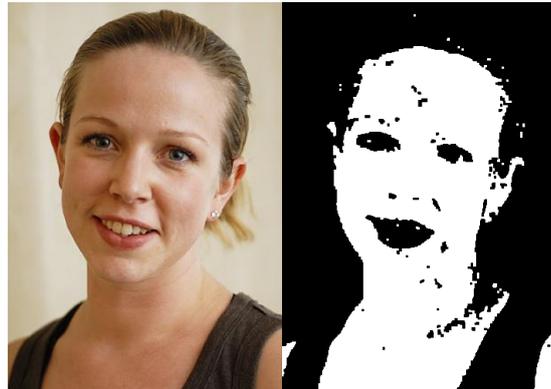


Figure 9. The result of improved skin-detection

Hence, the threshold standard for Y component is added to the back projection to search in a certain scope. Thus, the missing rate will not increase, while the false positive rate will decrease. The specific processes are as follows:

Step 1. A sample histogram model of Cb-Cr component is established with the N extracted skin color pixels, $P_i(Y, Cb, Cr)$ ($i=1,2, \dots, n$).

Step 2. The Y component is introduced. The sum of each Y component corresponding to m pixels $P_{bin-j}(Y, Cb, Cr)$ ($j=1,2, \dots, m$) in each bin of the histogram, and the mean are calculated:

$$E_{bin-y} = \frac{\sum_{j=1}^m P_{bin-j}(Y)}{m} \quad (7)$$

Where $P_{bin-j}(Y)$ is the value of Y component of the pixel $P_{bin-j}(Y, Cb, Cr)$ in each bin in the histogram.

Step 3. Meanwhile, the histogram of Cb-Cr component is established for all pixels in the test image.

Step 4. For the bin of the histogram where the pixel $P_{(w,h)}$ of the test image is located, the value of the corresponding bin is inquired for the pixel in the sample histogram model. Besides, the Y component of the pixel to be checked is compared with the corresponding E_{bin-y} in its bin. If there is $|Y - E_{bin-y}| < \beta$, then

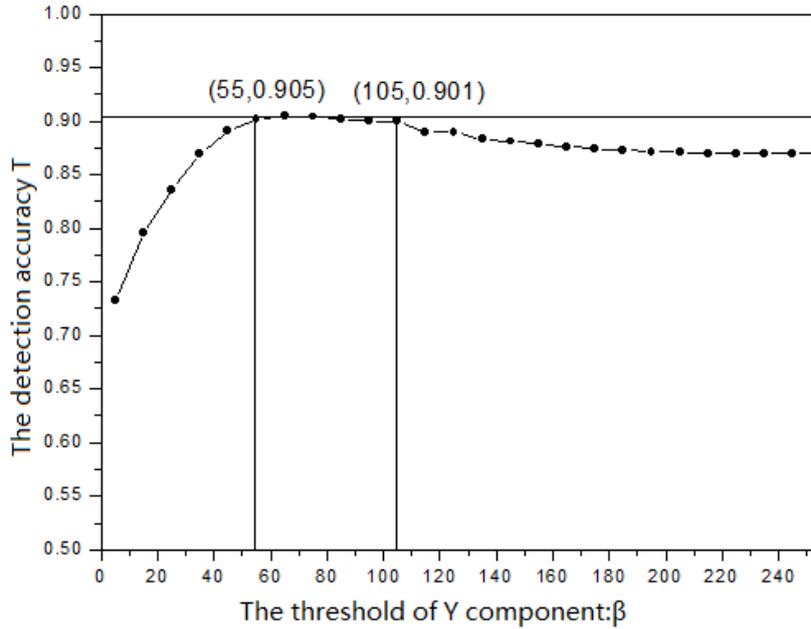


Figure 10. The relationship between the accuracy and threshold in sample data

the value of bin does not change; otherwise, it is changed to 0.

Step 5. The value of the pixel $P_{(w,h)}$ in the original image is replaced by the inquired value of bin, which is the process of back projection.

Step 6. The above process of back projection is applied to each pixel in the test image. The result is the skin color probability obtained after the projection of test image according to the sample skin color.

2.2.2 Result of Improved Algorithm

Taking Figure 7(a) as an example, the relationship between the detection accuracy T (formula 6) and the threshold β , and the effect of skin color detection can be obtained using the improved algorithm, as shown in Figure 8. In the figure, Y axis denotes the ratio of the sum of correct judgment of skin color and non-skin color to the total number of pixels in the test image, i.e., the detection accuracy T_{Accuracy} .

$$T_{\text{Accuracy}} = \frac{\text{The correct number of pixels}}{\text{The total number of pixels}} \quad (8)$$

It can be seen that with the changing of β , the accuracy increases and then decreases. This conforms to the regularity above: the Y component and the mean $E_{\text{bin-y}}$ are the same when $\beta = 0$, and the missing rate is higher; when $\beta = 255$, Y component is not considered, and the false positive rate is higher. Therefore, when the threshold β is 55, the detection effect is the best, as shown in Figure 9. The threshold of Y component for each test image is different. Hence, the best threshold should be found by the training with a number of images.

3 EXPERIMENTS AND ANALYSIS ON RESULTS

3.1 Data set for Experiment

The Pratheepan [11] data set of human skin color was used in this study. First, 12 human images were chosen from the 32 images in the dataset to train and obtain the optimal threshold of Y component β described in Section 1.2. The other 20 images were used as the test set to validate the result of training. The results were compared with those obtained with the traditional parametric method [1], self-learning skin color detection method (integration method) [11] and the luminance-based adaptive method [5].

3.2 Experimental Steps and Analysis

3.2.1 Training of the Threshold of Y component and the Validation

The 12 sample images for training contained 2,459,261 pixels, and the skin pixels and non-skin pixels are annotated manually. First, the method described in Section 2.1 is adopted to detect the face in each image and the extract sample skin pixels. Then, the skin pixels extracted from each image are utilized for the projection of all pixels in the image with the method described in Section 2.2, and the projected pixels are compared with the annotated results. As a result, the relationship between detection accuracy T and the threshold β is obtained as shown in Figure 10. It can be seen that the detection effect is optimal when $\beta \in [55, 105]$.

To validate the training results, the test set is used, and the method is the same with that of training pro-

cess. Only the judgment standard changes to accuracy $T_{\text{Precision}}$ (7), recall rate T_{Recall} (8), and F value (9). The 20 test samples contained 3,454,050 pixels, 793,581 of which were skin pixels. The detection results with different thresholds of the Y component are shown as Figure 11.

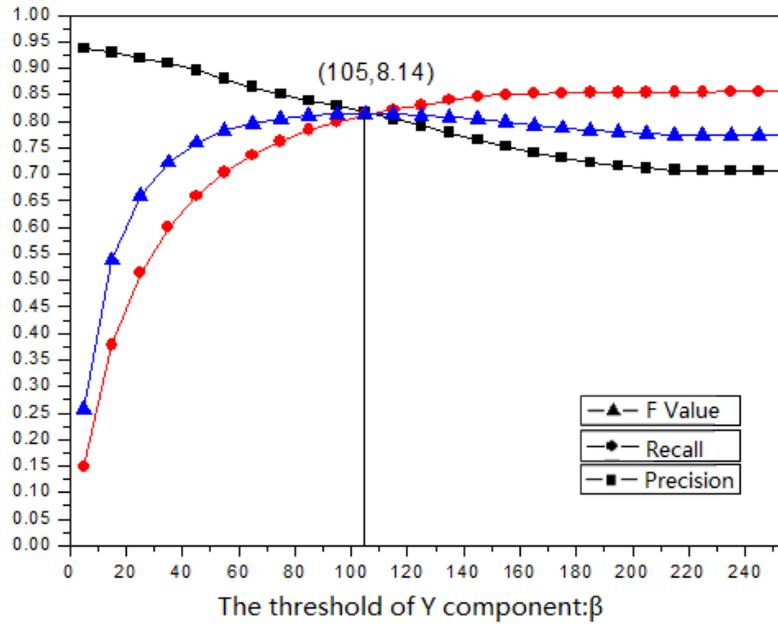


Figure 11. The test result in different threshold of Y weight



Figure 12. (a)The original image,(b)The result of traditional parametric method,(c) The result of integration method, (d) The result of this method

Table 1. The result of different approaches

The Method of Skin Detection	Accuracy	Recall	Precision	F Value
Parametric Method[1]	0.441	0.894	0.715	0.591
Integration Method[11]	0.689	0.655	0.877	0.676
Adaptive Method [5]	0.694	0.813	0.903	0.749
This Method	0.845	0.774	0.916	0.807

$$T_{\text{Precision}} = \frac{\text{The correct number of skin pixels}}{\text{The total number of detection pixels}} \quad (9)$$

$$T_{\text{Recall}} = \frac{\text{The correct number of skin pixels}}{\text{The total number of skin pixels}} \quad (10)$$

$$F_{\text{Measure}} = \frac{T_{\text{Precision}} * T_{\text{Recall}} * 2}{T_{\text{Precision}} + T_{\text{Recall}}} \quad (11)$$

It can be seen from the results that when the threshold of Y component $\beta = 105$, the judgment standard, F value, is maximal, which means the optimal balance between recall rate and accuracy. Moreover, the result conforms to the training result of the first step, where $\beta \in [55, 105]$.

3.2.2 Contrast Experiment

The parametric method, integration method and the luminance-based adaptive method were used for experiments with 20 images as the training, and the result was compared with that of the proposed method. Since the training result for the threshold of Y component β was [55, 105], $\beta = 80$ was set during the test process. The detection results are shown in Figure 12, and the parameter results are shown in Table 1.

It can be seen from the contrast of detection effect that although the parametric method has the worst performance on background judgment, the highest recall rate in Table 1 indicates the largest number of skin pixels detected. This means that this method is suitable for large-scale detection, while the effect is not satisfactory for some environments. As for the method proposed in this study, the processing of the background and the influence of luminance are both satisfying, with remarkable improvement in both accuracy and recall rate compared with the integration method. Although the recall rate of the online luminance-based adaptive method is superior to the proposed method, the accuracy of our method is higher, and the final outcome of the comprehensive judgment standard, F value, is the highest.

4 CONCLUSION

In this study, an adaptive skin color detection method is proposed by combining the color component and luminance component. After detecting human face, the sample pixels are not extracted based on the skin color features. Instead, by combining facial distribution feature and the improved local adaptive binarization method, the completeness of sample extraction is

guaranteed. Then, with the histogram backprojection algorithm, the interference of background pixels having similar color as the skin can be eliminated effectively. The accuracy of detection is ensured, while the false positive rate is also low. Meanwhile, the influence of luminance on skin color is also taken into consideration, which further increases the accuracy.

REFERENCES

- [1] L.J. Duan, G.Q. Cui, W. Gao, & H.M. Zhang. 2002. Adult Image Detection Method Base-on Skin Color Model and Support Vector Machine, ACCV. Melbourne, Australia, pp: 23-25, January.
- [2] J. Kovac, P. Peer, & F. Solina. 2003. *Human Skin Colour Clustering for Face Detection*, EUROCON Slovenia, pp: 144-148.
- [3] S. Bianco. & R. Schettini. 2014. Adaptive Color Constancy Using Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 36(8): 1505-1518.
- [4] J. J. Fang, Z. B. Li, & Y. Jiang. 2013. Human Skin Color Region Segmentation Based On Adaptive Model, *Journal of Computer-Aided Design & Computer Graphics*, 25: 229 - 234.
- [5] R. D. Wang. & Y. Yin. 2014. Luminance Adaptive Method for Skin Area Detection Based on Online Learning. *Journal of System Simulation*, 26: 2121-2125.
- [6] B. Yan. & L. Z. Liang. 2014. Improved Color Model Used in Face Detection, *Computer Simulation*, 31: 354-357.
- [7] Y. Zhu. 2012. Face Detection Method Based on Multi-feature Fusion in YCbCr Color Space, *5th International Congress on Image and Signal Processing, Chongqing, China*. USA: IEEE, pp: 1249-1252.
- [8] R. Hsu, M. Abdel-Mottaleb, & A. Jain. 2002. Face Detection in Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (S0162-8828), pp: 696-706.
- [9] Z. X. Chen. & C. Y. Liu. 2013. Fast face detection algorithm based on improved skin-color model, *Arabian Journal for Science and Engineering*, 38: 629-635.
- [10] Kim C, Choi S I. & Turk M. 2012. A new biased discriminate analysis using composite vectors for eye detection, *IEEE Transactions on Systems Man and Cybernetics: Part B-Cybernetics*, 42(4): 1095-1106.
- [11] W. Tan, C. Chan, P. Yogarajah, & J. Condell. 2012. A Fusion Approach for Efficient Human Skin Detection. *IEEE Transactions on Industrial Informatics* (S1551-3203), 8:138-147.