Chapter 12
Hand Segmentation Based on Skin Tone and Motion Detection with Complex Backgrounds

Xintao Li, Can Tang, Chun Gong, Sheng Cheng and Jianwei Zhang

Abstract Hand Segmentation is the first problem need to be solved in hand recognition system. Currently, most hand gesture recognition system is based on simple background, or requests the recognizer on glove in special color, which gives human–computer interaction some restrictions. This paper researches the gesture segmentation technology based on complex backgrounds, and gives a method combined with skin tone detection and motion detection. By experiments on the images captured by home security robot, this method can get accurate hand segmentation of all the images. This paper lays the foundation of gesture recognition on the home security robots.

Keywords Complex background · Hand segmentation · Skin color · Motion detection · Home security robot

12.1 Introduction

Gesture is a natural and intuitive interpersonal communication mode, therefore, in the field of human–computer interaction, Gesture recognition is the hot research topic, gesture recognition based on sequences (images) is the indispensable key technology of the new generation of human–computer interaction. Realize gesture recognition system need to solve the three important problems [1]: gesture segmentation, gesture analysis and gesture recognition. With the influence of complex of background and environment light, in the gesture recognition method based on monocular vision, how to division out gesture region is always a difficulty, many researchers used the method of limiting the gesture image, for example, use the

X. Li (✉) · C. Tang · C. Gong · S. Cheng · J. Zhang
Intelligent Robot Engineering Lab of Kunshan ITRI, 6F ITRI Building, No.1699, Weicheng South Road, Kunshan, Jiangsu, China
e-mail: lixt@ksitri.com

pure black or white wall, simplified background by dressed in the dark black clothing, or require people to wear special color gloves for outstanding hands area, etc. however, These methods increased the limitation of human–computer interaction, destroy the system availability and user-friendliness.

This paper mainly studies the gesture segmentation method within complex background, it segments hand Based on Skin Tone and Motion Detection. First, segment skin areas from complex background use skin color model. Then, get the moving regions by motion detection and filter the still skin areas in the background by mask the skin areas and moving regions. Last, get the accurate hand area by masking the moving hand areas and skin areas.

12.2 Skin Segmentation

The purpose of skin segmentation [2] is to separate the skin areas from the complex background, skin segmentation need to select appropriate color space and establish skin model. This paper uses the Gaussian skin model based on *YCbCr* color space.

YCbCr color space [3] can separate the luminance and chrominance of the image, *Y* component indicate the brightness of the pixel, *Cb* and *Cr* components called chrominance, *Cb* indicates blue component, *Cr* indicates red component, Color in this color space can be gathered in a very small range. *YCbCr* color space can full disclosure skin of body, and can maximum eliminate the influence of brightness, so reduce the number of dimensions of color space and reduce the computational complexity. We usually need to convert the *RGB* color space to *YCbCr* color space, the transformation formula as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & 0.331 & 0.500 \\ 0.500 & -0.419 & 0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{12.1}$$

Gaussian model [4] mainly use the principles of statistics, it believes random samples which conform to the normal distribution also meet Gaussian distribution such as skin color. The mathematical expression of Gaussian distribution is simple, intuitive, and is a normal model which research deeper in principle of Statistics. Gaussian model constitute a continuous data information by calculating the probability of pixel value and get a probability graph of skin color, then complete color confirmation by the probability of skin color. Gaussian can express as $N(m, C)$, m is mean value, C is covariance matrix.

$$m = E\{x\}, \ x = (Cr, Cb)^T \tag{12.2}$$

$$C = E\{(x - m)(x - m)^T\} \tag{12.3}$$

By the experimental statistics, Mean and covariance matrix respectively as:

$$m = (150.3179, \ 117.1057)^T \tag{12.4}$$

$$C = \begin{bmatrix} 250.2594 & 18.2077 \\ 18.2077 & 149.6103 \end{bmatrix} \tag{12.5}$$

By the Gaussian skin color model establish in advance, the probability of any pixel belongs to the skin can be calculated by the following formula:

$$P(Cr, Cb) = exp\left[-0.5(x - m)^T C^{-1}(x - m)\right] \tag{12.6}$$

Compute the skin color likelihood of all the pixels in the detected image, and get the maximum of skin color likelihood, then use the skin color likelihood of all the pixel divide the maximum skin color likelihood, we get the probability of the pixel belongs to the skin color. The image composed by skin color probability of all the pixel is called color likelihood image, in the color likelihood image, we set a threshold, when the pixel value greater than the threshold, we can confirm the pixel is skin pixel, then we can get the segmentation skin image. At last, corrode and dilate skin color detection result image, some skin like small areas can be eliminated. The results are shown as Fig. 12.1.

12.3 Motion Detection

The purpose of motion detection is to extract the changed area in the sequence images from the background. The effective segmentation of the moving regions is essential to the later processing of target classification, tracking and behavior understanding. However, because of the dynamic changes of background image, for example, influence by the weather, illumination and shadow, make the motion detection to be a very difficult work. Commonly used motion detection methods are Background Subtraction [5, 6], Temporal Difference [7] and Optical Flow [8].

In the hands waved process, hands is a motion area, therefore, we can eliminate the disturbance of color like regions in the static background by motion information. Based on the efficiency of algorithm consideration, this paper uses the method of Temporal Difference for motion detection. Temporal Difference (also

Fig. 12.1 Skin segmentation **a** original image, **b** result of skin segmentation, **c** corrosion and expansion results

called Adjacent frame Difference) method extract the moving area by temporal difference based on pixel in continuous image sequence and threshold. Temporal Difference has strong adaptability for dynamic environment. The shortcoming of this method is can't detect the overlap part of the moving object, caused incomplete of the moving object, and produce empty in the internal of the moving object. In order to solve the problems, this paper selects the discontinuous and frames which have obvious movement for difference, as shown in Fig. 12.2.

From the results above, we can see that motion detection not only detected the moving hand region, but also detected the body and head movement, these movement are not we need, so in order to exactly segment moving hand region, we need to further remove the useless areas.

12.4 Hand Segmentation

To perform the mission of hand segmentation three phases are introduced.

12.4.1 Skin Color Mask

Mask the motion detect result and skin color detect result, here we use and operation, by this step, we can effectively remove the moving non-skin regions, include body and other moving objects in the image, get the moving skin color regions. The results of this step are shown in Figs. 12.3 and 12.4.

12.4.2 Motion Mask

From the results of Sect. 12.4.1, we can see that, the results of masked motion regions and skin color regions contain the moving face regions, so we need to further eliminate these regions. By analyzing the results of the first step, we get the
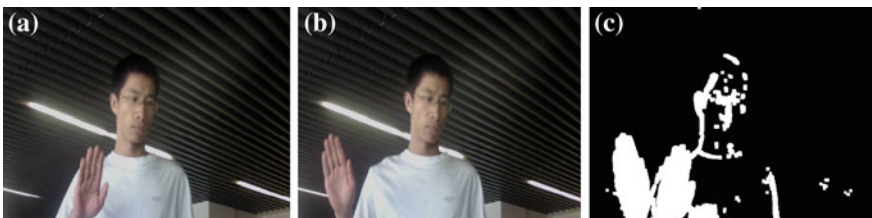

Fig. 12.2 Motion detection **a** frame 1, **b** frame 10, **c** motion detection results

Fig. 12.3 Skin color mask (1) **a** motion detection result, **b** skin detection result (1), **c** mask result (1)
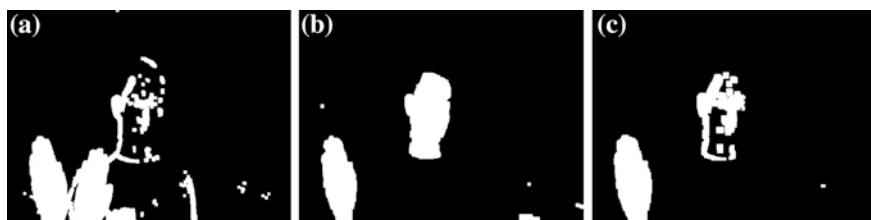

Fig. 12.4 Skin color mask (2) **a** motion detection result, **b** skin detection result (2) **c** mask result (2)

conclusion that again masks the moving skin regions can eliminate the face skin regions. The result of this step is shown as Fig. 12.5.

12.4.3 Hand Region Extraction

From the results of Sect. 12.4.2, we can see that, skin like regions in the background and face regions are completely eliminated. But both of the two hands are saved. Obviously, again masks the results of motion mask and skin detection, we can get the signal hand region in the current image. The results of this step are shown as Figs. 12.6 and 12.7.

Fig. 12.5 Motion mask **a** skin color mask result (1), **b** skin color mask result (2), **c** motion mask result

Fig. 12.6 Hand region extraction (1) **a** skin color detect result (1), **b** motion mask result, **c** hand segmentation result (1)

Fig. 12.7 Hand region extraction (2) **a** skin color detect result (2), **b** motion mask result, **c** hand segmentation result (2)

12.5 Experiments

In order to check the validity of the method proposed in this paper, we use the surveillance camera in the home security robot shooting different environment of videos, select different video frames for experiments, at the same time, use this method to hand wave direction recognition of the home security robot. From the results, this method achieve recognition rate of 90 %, which greatly improving the interactive performance of the home security robot. Some hand region segmentation results are shown as Fig. 12.8a–e are original images, f–j are corresponding hand region segmentation results.
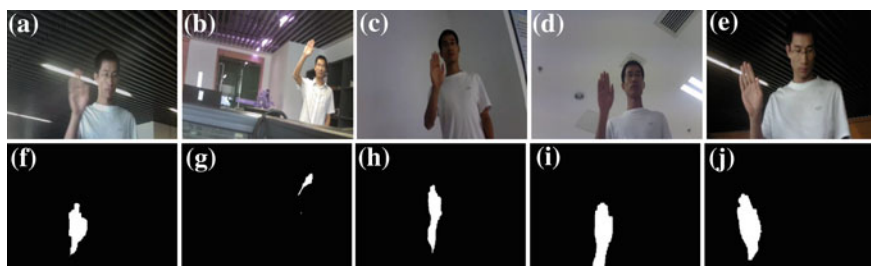

Fig. 12.8 Hand region segmentation results

12.6 Conclusion

This paper aims at the difficulties of hand region segmentation with complex background, uses a method combined with Skin Tone and Motion Detection, realize accurate hand region segmentation from coarse to fine. First, use the Gaussian skin color model based on *YCbCr* color space to detect skin region. Then, detect motion from two discontinuous image frames, eliminate non-skin moving regions by skin mask and eliminate face skin regions by motion mask. In the end, again mask the results of motion mask and skin regions, accurately segment out the hand region in the current image. Through the results of the experiments, this method can better segment hand regions under different complex backgrounds, achieve accurate hand wave direction recognition, greatly enhance interactive performance of the home security robot.

References

1. Chen Y, Zhang Y (2009) Research on human-robot interaction technique based on hand gesture recognition. Robot 31(4):351–356 (in Chinese)
2. Chen D, Liu Z (2006) A survey of skin color detection. Chin J Comput 29(2):194–203 (in Chinese)
3. Zheng N (1998) Computer vision and pattern recognition. National Defence Industry Press, Beijing
4. Xu Z, Zhu M (2007) Color-based skin detection: a survey. J Image Graph 12(3):377–388 (in Chinese)
5. McKenna SJ, Jabri S, Duric Z (2000) Tracking groups of people. Comput Vis Image Underst 80(1):42–56
6. Haritaoglu I, Harwook D (2000) W4: real-time surveillance of people and their activities. IEEE Trans Pattern Mach Intell 22(8):809–830
7. Lipton A, Fujiyoshi H, Patil R (1998) Moving target classification and tracking from real-time video. In: Proceedings of IEEE workshop an application of computer vision, Princeton, NJ, pp 8–14
8. Meyer D, Denzler J, Niemann H (1997) Model based extraction of articulated objects in image sequence for gait analysis. In: IEEE international conference on image processing, vol 3, pp 78–81