Research on gesture recognition preprocessing technology based on skin color detection

Wenquan Tang¹, Wu Xu¹*, Xing Guo¹, Cong Wen¹, Bing Zhou²

¹School of Electric and informative Engineering, Yunnan Minzu University, Kunming, Yunnan, 650504, China
²Yunnan Agricultural University, Kunming, Yunnan, 650500, China
*Corresponding author’s e-mail: 1634298994@qq.com

Abstract. In the process of gesture recognition, human hand detection based on skin color is a method with high recognition rate. When the acquired gesture information is converted into digital images that can be processed by computer, the gesture images will be interfered and affected by various factors in the process of generation, transmission or transformation, and the quality of the gesture images will be distorted to different degrees due to noise, so the image needs to be preprocessed first. The process of image preprocessing includes image binarization and image smoothing. In this paper, the smoothing technology and binarization method are respectively studied and compared in experiments, providing effective data samples for gesture segmentation and feature extraction in gesture recognition.

1. Introduction

In the era of rapid development of science and technology, image processing technology has been widely developed in various fields such as economy, aerospace and military, and has begun to facilitate national life. The research of image preprocessing algorithm is also very important. Image preprocessing algorithm is the image processing before the deep processing of the image, to improve the transmission efficiency of image information, reduce the noise interference in the image propagation process, improve the image definition and simplify the image propagation process.

Through digital image preprocessing technology, we can greatly compress the image, so as to realize high speed and accurate transmission and preservation of the image, greatly improving the transmission and preservation process[1]. In practical production and application, digital image preprocessing technology is subject to many limitations. In digital image preprocessing, a large amount of data and algorithms need to be processed, and a lot of time is spent. Meanwhile, the power consumption, real-time performance, occupied area and production cost of the platform also become important criteria to measure the reliability of the platform.

The purpose of preprocessing is to remove the noise and enhance the useful information in the image. Image preprocessing is a filtering process of images, which eliminates interference and retains the parts needed, while filtering out the parts not needed. The main work of the gesture image preprocessing part is to binarize the original gesture image according to the skin color characteristics of the human body, and denoise the binaries of the gesture image by image enhancement to get the binaries without noise, and then extract the edge of the binaries to get the outline of the gesture. This paper introduces the preprocessing of gesture image from three parts: skin color feature, noise processing and edge extraction.
2. Human skin color characteristics

Skin color detection is used in human hand detection, mainly because color is one of the most prominent features, using color to detect human hand is a natural idea. However, the color of the surface of the human hand is influenced by a variety of factors, including individual differences in human beings, the light source and the Angle of illumination, resulting in highlights and shadows. Therefore, the task of dividing human hands by color becomes a very difficult problem. Based on human skin color detection, the key is to choose the right skin color model and skin color clustering [2]. Therefore, this paper uses these two color spaces for skin color segmentation.

2.1. Color balance

Due to the influence of light and shade, the acquired gesture image will have color deviation when the light is not ideal, which will lead to a large noise in the gesture data. Therefore, Gray World Theory (GWT), a color balance algorithm, is introduced for color correction. The GWT algorithm is expressed as follows:

\[
R = R \times \left[ \frac{K}{R_{AVG}} \right] \\
G = G \times \left[ \frac{K}{G_{AVG}} \right] \\
B = B \times \left[ \frac{K}{B_{AVG}} \right]
\]

\[
K = \frac{R_{AVG} + G_{AVG} + B_{AVG}}{3}
\]

RGB is the color of the red, green and blue channels. This standard includes almost all colors that can be perceived by human eyes, and it is one of the most widely used color systems.

2.2. Gesture region extraction based on Gaussian skin tone modeling

Because the background of gesture image is complex, and the brightness of skin color will change under different lighting changes, it is necessary to adopt a reliable skin color model to detect the gesture area. The results showed that differences in brightness were much smaller than differences in chroma [3]. Therefore, Gaussian distribution is used to model skin color in YCbCr space, and the probability value of skin color of each point in the image can be calculated, so as to segment gesture area. The calculation formula of skin color modeling based on Gaussian distribution is shown in equation (5).

\[
P(C_b, C_r) = \exp \left\{ -0.5(x - m)^T C^{-1}(x - m) \right\}
\]

Among it, \( x = \left(C_b, C_r \right)^T \), \( m = E(x) \), \( C = E \left\{ (x - m)(x - m)^T \right\} \)

A complete skin color probability distribution matrix can be established by calculating the probability value \( P \) of skin color belonging to each pixel in the image. By introducing OTSU algorithm, it means that the threshold value \( T \) should maximize the variance between target and background. The image is divided into two parts: background and target. The OTSU method is adopted to binarization the adaptive threshold of the color probability matrix. In the binarization image, when pixel value is 1, it means that this area is a skin color point. Otherwise, the pixel value is 0, which means that the region is a non-skin color point [4].
3. Filtering processing

The aforementioned skin color characteristics are determined by human and have certain subjective factors, so the binarization conditions are neither sufficient nor necessary. Therefore, the binarization binaries are mainly influenced by salt-pepper noise. Since median filtering can eliminate image noise and retain image details, it is not enough to use median filtering for processing of larger noise areas in binary images [5]. In this paper, median filtering is used to denoise binary graph of gesture region.

3.1. Binarization

Binarization is to set the grayscale value of the pixel on the image to 0 or 255, that is, to display the entire simulation image with different color effects, and set the color to black and white. When image binarization is conducive to further processing of the image, the collection property of the image is only related to the location of the point with pixel value of 0 or 255, and has nothing to do with pixel value, making the processing easier and the amount of data processing and compression small [6].

Suppose the gray value range of image f(x, y) is set as [a, b], and the color threshold of image simulation is set as t, then the calculation formula of binarization of the image is:

\[
g(x, y) = \begin{cases} 
0, & f(x, y) < t \\
1, & f(x, y) \geq t 
\end{cases}
\]  

(6)

Where g(x, y) is the binarization image obtained, and the threshold t selected is different, so is the binarization image g(x, y) obtained. Therefore, to achieve the ideal segmentation effect, it is very important to select the appropriate threshold t. The binarization graph obtained by gesture detection is shown in figure 1:

![Figure 1. Gesture binary graph](image)

3.2. Image smoothing processing

The median filter and Laplacian operator are used to process the image data. The Laplace operator will further enhance the noise during the detection process, which will affect the definition of the image, resulting in the weak simulation effect. Therefore, when the Laplacian operator optimizes the detection of images, it needs to be smoothed to reduce the noise interference and improve the image processing effect. The advantage of median filter is that it has advantages in eliminating long tail noise and white noise, but it should not be used when there are many details in the image, such as points, lines and spikes. Compared with mean filtering, median filtering is more complicated, but the picture is clearer [7].

In one-dimensional environment, median filtering is to replace the value of a pixel with the median gray value of its neighboring pixels. If the original signal sequence is represented, and the sequence is from large to small, then the median value can be expressed as:

\[
y = \text{Median}\{a_1, a_2, a_3, \cdots, a_n\} = \begin{cases} 
a \left(\frac{n+1}{2}\right), & n \text{ is odd number} \\
\frac{a_n + a_{n+1}}{2}, & n \text{ is even number} 
\end{cases}
\]  

(7)

Where: Zmin= the minimum gray value in Sxy, Zmax= the maximum gray value in Sxy, Zmed=...
the median gray value in Sxy, Zxy= the gray value at coordinate (x, y), and Smax= the maximum size allowed in Sxy. Compared with the median filtering, the improved median filtering algorithm can completely reduce noise interference and deal with impulse noise with higher probability, such as salt-pepper noise.

Since the Laplace operator is a second derivative, it will produce a highly variable intersection at the boundary of the image. In addition, because noise points have certain influence on edge detection, Laplace operator is a good edge detector [8]. For the image without noise and with steep edge, it can be detected by Laplace operator. In the Laplacian operator, we take the first derivative of the pixel value of a point, and then take the second derivative. The second derivative curve can well reflect the inflexion point of the original curve, which is a point on the boundary between the image and the background image. The zeros of the second derivatives don't just appear on the edges, they can also appear in nonsensate positions, but you can filter them out by doing this. In image processing, the second derivative is calculated through Laplace template, and its discrete form is defined as follows:

\[
\nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}
\]

\[
= f(x+1, y) + f(x-1, y) - 4f(x, y) + f(x+1, y+1) + f(x, y-1)
\]

Laplace operator can also be expressed in the form of template. The corresponding template of the above equation is shown in figure 2 (a) and 2 (b):

| 0 | 1 | 0 |
|---|---|---|
| 1 | -4 | 1 |
| 0 | 1 | 0 |

Figure 2 (a). Laplace operation template

| 1 | 1 | 1 |
|---|---|---|
| 1 | -8 | 1 |
| 1 | 1 | 1 |

Figure 2 (b). Laplace operation extension template

It is easy to see from the template form that if there is a point in a darker area of the image that is much brighter than the surrounding area, then the point in the image will be brighter after using the Laplace transform. General enhancement techniques are difficult for steep edges and slowly changing edges to accurately locate their edge lines. Therefore, it is very suitable for the purpose of highlighting the isolated points, isolated lines or line endpoints in the image.

The gesture diagram obtained after median filtering and Laplace image sharpening is shown in figure 3:

Figure 3. Preprocessing result image

4. Conclusion

After the preprocessing of the image obtained by skin color detection, the redundant information and interference noise of the image can be well removed, the integrity of the image information can be guaranteed, and the success rate of data transmission can be improved, which lays a solid foundation for the further processing of digital images. This paper studies a series of classical algorithms in image preprocessing, including image binarization, image median filtering and image Laplacian operator edge detection. It can be seen from the results of simulation detection that the image preprocessing results can overcome many defects, including partial blur and large noise interference in the original image. The results obtained by processing are more clear and noiseless, providing effective data
samples for gesture segmentation and feature extraction in gesture recognition.

Acknowledgments
This paper is supported by the national ethnic research foundation project "research on the digital protection and development of multi-cultural diversity of ethnic minorities under the big data background" (2019-GMD-056).

References
[1] Hu Y. Overview of gesture recognition technology [J]. China science and technology information. 2005, 1(2): 41-42.
[2] Wang H., Shi N., Zhao H. et al.Image preprocessing technology of improved median filtering method [J].Computer system applications,2015(5):237-240.
[3] Guo H., Xu L., Zhao H. et al.A morphological filter for suppressing speckle noise of sonar image [M].Chinese journal of instrumentation,2015,36(3):654-660.
[4] Li B., peng T., Peng B. Intelligent image processing technology [M]. Beijing: electronic industry shanban, 2004: 203-209.321-333.
[5] Tao L., Peng Z., Xu G.. Skin color characteristics of human body [J]. Journal of software.2001, 12(7): 1032-1041.
[6] Liang N..Research on image preprocessing technology in gesture recognition [J].Wireless interconnection technology, 2015(1) : 161-162.
[7] Zhou G., Liu huizhong. Implementation of edge detection algorithm [J].Computer system applications, 2015,24(10):271-275.
[8] Xue J., Zong Y., Yang Z. Gesture recognition based on improved YCbCr space and multi-feature fusion [J]. Computer application and software.2016(01).