Application of static gesture segmentation based on an improved canny operator

Shenjian Gong\textsuperscript{1,2}, Guangqiang Li\textsuperscript{2}, Yongjiu Zhang\textsuperscript{3}, Changdi Li\textsuperscript{2}, Lei Yu\textsuperscript{1,4,2}  
\textsuperscript{1}Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing, People’s Republic of China  
\textsuperscript{2}School of Mechanical and Electric Engineering, Soochow University, Suzhou, People’s Republic of China  
\textsuperscript{3}Jiangsu Province Special Equipment Safety Supervision Inspection Institute Branch of Wujiang, People’s Republic of China  
\textsuperscript{4}Collaborative Innovation Center of Industrial Energy-Saving and Power Quality Control (Anhui University), Hefei, People’s Republic of China  
\textsuperscript{✉}E-mail: yu_lei@suda.edu.cn

Abstract: In order to meet the requirements of gesture extraction in static gesture segmentation, this study proposes to improve the traditional Canny operator by using a combined filtering method and an adaptive threshold algorithm. The original image is firstly denoised by bilateral filtering, and then the elliptical skin model is used to select the color domain, and the domain is divided into blocks. The hand is separated according to the block domain feature. Finally, an improved Canny operator with an adaptive threshold is used to extract the interactive hand contour. Experiments show that it has high accuracy in simple static gesture segmentation.

1 Introduction

Vision-based gesture recognition is the abstraction of various information features of the handle from gestures, which are widely used in various human–computer interaction systems [1–14]. Among them, the most common way to extract gestures is to extract the contour of the hand to describe the characteristics of the hand posture, shape position, and so on. With the rapid development of machine vision technology, segmentation, and extraction gesture more with a layer of stairs.

The edge-detection technology of static hand is one of the research hotspots of image processing technology. In terms of gesture segmentation, the early use of auxiliary colour marking method [15], there is the thermal infrared capture method [16, 17], the method is simple, but the external conditions are limited. There are also geometric feature-based segmentation methods which increase the complexity of the algorithm due to morphological variability. After separating the hand image, the contour extraction separated [18–20]. The hand image is used and the image is increase the complexity of the algorithm due to morphological edge detection than the above operator, but there are also problems in that it is difficult to filter out impulse noise and the contour threshold cannot be automatically set.

In this paper, the elliptical colour model based on the YCrCb space is used in gesture segmentation [7]. The colour domain is first selected by this model, and the colour domain is divided into blocks. The hand is separated according to the relative characteristics of the block, and the hand image is extracted and separated [18–20]. The hand image is used and the image is extracted by the adaptive threshold method to calculate the upper and lower thresholds for Canny gesture contour extraction.

2 Canny operator and gesture segmentation

Actually, the gesture is reflected by the contour of the hand edge. It is very practical and popular to use the Canny edge-detection algorithm to extract the hand contour. After using the elliptical colour model and the block segmentation to filter out the required partial images, the basic idea of the traditional Canny algorithm is to detect whether the point is the local gradient maximum point or the minimum point by detecting the gradient value of the image gradient. This determines the contour information of the image, which in turn extracts the contour of the hand. The Canny operator was originally proposed by John F. Canny, which is a multi-step composition algorithm.

2.1 Canny algorithm principle steps

2.1.1 Noise removal: Filtering out the noise before processing the image is a necessary step. In the Canny algorithm, the traditional Gaussian filtering method is selected. In the algorithm, a $5 \times 5$ Gaussian convolution kernel and a filtered image noise are used to achieve image smoothing. To prevent noise from being detected as an edge. The Gaussian kernel is calculated as follows [4]:

$$G_{x, y, \sigma} = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{x^2 + y^2}{2\sigma^2} \right)$$  \hspace{1cm} (1)

2.1.2 Calculating the image gradient: After the image is denoised, the Sobel operator is used to take the first order of the horizontal and vertical directions, respectively, and the image gradients in both directions are $G_x$ and $G_y$. According to the bidirectional gradient, the two-dimensional real gradient direction can be obtained:

$$A(i, j) = \frac{1}{2} \sqrt{G_x(i, j)^2 + G_y(i, j)^2}$$ \hspace{1cm} (2)

$$\theta(i, j) = \arctan \left( \frac{G_x(i, j)}{G_y(i, j)} \right)$$  \hspace{1cm} (3)

The element $A(i, j)$ is reflected in the maximum change at the pixel of the coordinates $(i, j)$, and the element $\theta(i, j)$ is reflected in the direction of the maximum edge change at the pixel of the coordinates $(i, j)$. In order to facilitate the storage direction, the gradient directions are grouped into four categories, vertical, horizontal, and double diagonal, and the gradient direction is likely to be perpendicular to the boundary.

2.1.3 Non-maximum suppression: Although the gradient of each pixel is obtained, no boundary is found, so it is necessary to filter out the points on the true edge. The method is to scan each pixel in turn and check whether the element has the largest change
among the points with the same gradient direction around. The principle is shown in Fig. 1.

As shown in Fig. 1, when scanning to point A, the change value of the surrounding B and C points in the gradient direction of point A is judged. If point A has the largest change value of the surrounding elements in this direction, the screen can be selected. Point A is the edge point. The previous direction classification can find the comparison points in the gradient direction faster.

2.1.4 Get edge: After non-maximum suppression, a continuous boundary of a strip is obtained. At this time, set two thresholds, minval (lower threshold) and maxval (upper threshold); the screening mode is shown in Fig. 2:

Firstly, the scatter points need to be removed, and the selected boundary has enough length. On this basis, it is assumed that there are two consecutive edges $\Theta$, At $\Theta$, there is a point A where the image gradient is higher than maxval, and even if the lowest point C is smaller than maxval but is at the same edge as A, it is considered to be valid. Conversely, there are no element points with gradients greater than maxval in $\Theta$, even if all points change above minval, they cannot be considered valid boundaries.

2.2 Improved filtering algorithm

The traditional Canny operator achieves image denoising by using Gaussian filtering in the initial filtering. However, this filtering method also weakens the edges and reduces the edge gradient while optimising the image. Traditional median filtering and mean filtering also have this problem [8–14].

To this end, this paper proposes the use of bilateral filtering for image denoising. Bilateral filtering is the simultaneous consideration of spatial domain information and numerical domain information of the pixel to be filtered. Therefore, it belongs to a joint filtering method, which uses spatial Gaussian weights and grey value similarity weights. The spatial Gaussian blur smooths the image with the same convolution. The similarity of the grey values ensures that only the pixels similar to the centre grey will affect the centre as the weight parameter in the convolution matrix, so the convolution corresponding to the pixel is dynamically changing. Therefore, the bilateral filtering can ensure the clarity of the boundary as much as possible, because the grey-level mutation is large at the boundary.

Its mathematical expression is

$$h(x) = k^{-1} \int_{0}^{\infty} w_{\sigma}(y) \phi_{\sigma}(f(y) - f(x)) f(y) dy$$  \hspace{1cm} (4)

$$k = \int_{0}^{\infty} w_{\sigma}(y) \phi_{\sigma}(f(y) - f(x)) dy$$  \hspace{1cm} (5)

In the formula, $\nu$ is the convolutional domain, $\sigma_{\nu}$ is the standard deviation of the spatial Gaussian function, $\sigma_{\nu}$ is the standard deviation of the grey-scale similarity Gaussian function, $y$ is the element in the convolution domain, $x$ is the desired point, $w_{\sigma}(y)$ is the spatial convolution kernel normalisation element. $\phi_{\sigma}$ is a normalisation matrix of $f(y) - f(x)$.

2.3 Improved threshold algorithm

The traditional Canny operator needs to manually select the upper and lower thresholds when the edge is finally obtained. The difference between the images is too large, and the value needs to be changed according to the specific image. This paper proposes the use of the adaptive threshold algorithm [4].

Suppose that the image has a total of $N$ points, and $P_{i}$ is the number of $i$ grey points. So it satisfies

$$N = \sum_{i=0}^{255} P_{i}$$  \hspace{1cm} (6)

Assuming that the threshold $K$ is taken, the gradation is divided into two categories $[0, k]$ and $[k + 1, 255]$, and the average gradation $E(X)$ satisfies the following formula:

$$E = \int_{0}^{255} iP_{i}$$  \hspace{1cm} (7)

Similarly, the mean values $E_{1}$ and $E_{2}$ of classes 1 and 2 exist.

The optimal threshold condition is that the variance between classes is the largest, and the formula is expressed as

$$\max \left( \left( \sum_{i=0}^{k} P_{i} \sum_{i=k+1}^{255} P(E_{1} - E_{2})^{2} \right) / N^{2} \right)$$  \hspace{1cm} (8)

Select $k/2$ as a small threshold to achieve an adaptive threshold.

2.4 Static gesture segmentation

2.4.1 Elliptical colour model: In the case of simple hand separation, this paper uses an elliptical skin colour model based on the YCrCb space. It was proposed by Rein et al. to map the RGB space to the YCrCb space:

$$Y = 0.299R + 0.587G + 0.114B$$  \hspace{1cm} (9)

$$Cr = -0.147R - 0.289G + 0.436B$$  \hspace{1cm} (10)

$$Cb = 0.615R - 0.515G - 0.100B$$  \hspace{1cm} (11)

It was discovered by scientists and through training samples that skin colour clustered in an ellipse on the Cr−Cb plane in the YCrCb space. The elliptical centre $Cy = 109.38$, $Cy = 152.02$, the long axis is 25.39, the short axis is 14.03, and the counter-clockwise rotation of the $x$-axis is $\theta = 2.53$ rad.

2.5 Block domain segmentation algorithm

This paper uses the segmentation method more suitable for the presence of the face. The image is traversed, and the skin colour detection part is divided into three large blocks of the face and the left and right hands. In this case, the screening threshold can be selected by the block area and the perimeter feature, and the dynamic upper and lower area thresholds take the largest area of $1/k_{1}$, $1/k_{2}$, the upper and lower perimeter thresholds take $1/k_{3}$, $1/k_{4}$.
of the maximum circumference. The hand separation block is obtained, and then the image contour extraction is performed using the improved Canny algorithm.

3 Simulation results

The simulation environment for this paper is Python 3.6, using the integrated library OpenCV3 as the image processing core.

Firstly, the YCrCb space-based elliptical colour model detection algorithm is used to extract the fish skin colour domain, and the hand skin colour domain is then obtained by block filtering and filtering out the small colour domain.

In this experiment, the traditional Canny operator and the improved Canny operator are compared based on the extracted part of the skin colour domain. The experimental results are shown in the following figure. Fig. 3 shows the original image and the original image with Gaussian noise. Fig. 4 shows the original image with Gaussian noise added after Gaussian filtering and bilateral filtering. The subsequent processing is based on the bilateral filtering. Fig. 5 shows the original skin region and its mask extracted using the elliptical skin colour model. Fig. 6 shows the separation block. The original hand colour block and its mask are obtained from the domain; Fig. 7 is the hand edge-detection map of the Canny operator with an improved adaptive threshold and the traditional Canny operator.

From Fig. 4, it can be found that, as mentioned above, Gaussian filtering is only used to consider the spatial weighting coefficient, and the degree of blurring is almost uniform everywhere, so the subsequent gradient detection cannot reflect the boundary variation more realistically. The bilateral filtering can maintain a clear contour with a large boundary gradient, and the accuracy and efficiency of edge detection are improved. From Fig. 5, using the elliptical colour model based on big data training, the majority of the colour in the picture can be extracted. Although the eyes and part of the hair are included, the error is within the acceptable range, and later screens it off. Fig. 6 is the result of segmentation selection in the colour block. By adaptive block domain segmentation, only the hand contour with the connected domain area is within a certain range is extracted. This method is more feasible for most positioning interaction modes. The final result of the experiment is shown in Fig. 7. The adaptive upper and lower threshold method makes the outline of the hand clear and can reflect most of the gesture information. However, the threshold selection of the traditional Canny operator self-set is likely to be inappropriate, which makes it difficult to remove the curved details of the finger and the large veins of the hand.

Experiments show that the use of bilateral filtering can be used for noise filtering of images that require edge extraction, and the result is better than traditional Gaussian filtering, which can avoid blurring of grey-scale abrupt edges, preserve the original image gradient around the hand contour, and its effect on the operation of the screening edge is reduced. The Canny operator with adaptive threshold can automatically solve the grey-scale boundary point according to the maximum variance method between grey classes in simple static gesture segmentation. Therefore, the maximum resolution of the image can be achieved when the edge is filtered, and the final edge is obtained. The edge with the highest contrast of the image has higher adaptability to different images, and the Canny operator has higher efficiency and accuracy in extracting the gesture contour. The combined use of the two guarantees the accuracy of the image while showing a relatively obvious contour.

4 Conclusion

There are many problems with the contour extraction algorithm used in traditional gesture segmentation. They are either simple but very sensitive to noise or have poor adaptive capabilities and more false edges like the traditional Canny algorithm. This paper proposes to improve the traditional Canny operator by using a combination filtering method and adaptive threshold algorithm. Firstly, use bilateral filtering on the image to highlight the edges when filtering out noise. Secondly, the hand region is extracted by skin colour detection and block domain segmentation. According to the overall colour gamut distribution of the image, an adaptive threshold algorithm is used to calculate the upper and lower thresholds for filtering the edges, which can reduce the false edges. Practice has proved that this algorithm has high efficiency and accuracy in simple static gesture segmentation. It is an edge-detection algorithm suitable for static gesture segmentation.

5 Acknowledgments

The work is supported by the National Natural Science Foundation of China (61873176); Natural Science Foundation of Jiangsu Province, China (BK20181433); Fundamental Research Funds for the Central Universities (30918014108); Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX18_0844); Open Fund for Collaborative Innovation Center of Industrial Energy-Saving and Power Quality Control, Anhui University (KFKT201806); Natural Science Fund for Colleges and Universities in Jiangsu Province (16KJB120005); The Key Technology Program of Suzhou, China (SYG201639). The authors would like to thank the referees for their constructive comments.
6 References


[18] Zhang, J., Chang, W., Wu, L.: ‘Edge detection based on general grey correlation and LoG operator’. Int. Conf. on Artificial Intelligence and Computational Intelligence, Hangzhou, China, 2010, pp. 480–483


This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)

546