ISSN: 2664-9640

Research on Gesture Recognition Technology Based on Convex Hull Algorithm

Miaomiao Chen [a], Dan Zheng [b], Xiaoqiang Reng [c]

School of Information Engineering, Southwest Jiaotong Hope College University, Chengdu, 610000, China;

[a] 979485861@qq.com, [b] 963675872 @qq.com, [c] 946551845 @qq.com

Abstract

In this paper, the problem of unsatisfactory hand extraction effect based on skin color is improved. The skin color segmentation method fused with HSV and YCrCb color space is used, and the convex hull algorithm is used for static gesture recognition. The experimental results show that the algorithm in this paper has strong stability in the recognition of static gestures in scenes with different light intensities.

Keywords

Gesture recognition, monocular camera, convex hull algorithm.

1. Introduction

At present, the research on gesture recognition based on monocular camera has been carried out for decades, and many results have been obtained. Domestic institutions engaged in research in this field include the State Key Laboratory of CAD&CG, Zhejiang University, and the Institute of Human-Computer Interaction and Media Integration, Tsinghua University. Foreign research in this field is earlier, and there are also many universities and institutions specializing in research in this field, such as the Massachusetts Institute of Technology Media Laboratory, and the Computer Department of Michigan State University. In the commercial market, this technology has been widely used. Tencent launched the "QQ Gesture Master" in 2011, which uses gesture recognition to control PPT playback. Konka also launched the country's first gesture interactive TV in 2012. Toshiba of Japan has added image recognition and gesture recognition technology to the Qosimo product line, and has developed notebook computers and TVs that can interact with gestures [1].

In gesture recognition, the segmentation of the hand area and the background area is the primary issue. At present, the mainstream gesture detection method is based on color information. The early detection is mainly based on RGB color value. However, this method is greatly affected by light. In response to this situation, researchers have proposed the use of multiple colors. Space scheme. For example, Liu Jun et al. used HSI color space for skin color segmentation, and the recognition rate was above 90% [2]. Another example is Qin Wenjun's use of YCbCr color space for skin color segmentation [3]. Different color spaces have different differences in separating the color components and light components of the skin itself. This difference determines the pros and cons of the segmentation method to a certain extent. From this, a variety of color space fusion detection methods have been derived, such as Zhang Huan , An Guocheng and others use the SV component of HSV space and the Cb component of YCrCb space for human body detection [4]. This algorithm can better detect the human body when the camera is fixed, the light changes slowly, and there is camouflage. In gesture recognition, there are mainly methods based on template matching and methods based on appearance features. Methods based on template matching usually use machine learning methods such as support vector machines and neural networks to recognize gestures. Kai-ping Feng et al. extract the

ISSN: 2664-9640

HOG of gestures and recognize them through support vector machines. The average recognition rate of the system can reach 90%. Above [5].

This paper studies the gesture recognition technology based on monocular camera. Research on hand region extraction under static gesture recognition, fusion of HSV and YCrCb color space to build a skin color model to improve the accuracy and stability of recognition, and use convex hull algorithm for static gesture recognition.

2. Theoretical basis of gesture recognition

2.1. Skin tone detection

Skin color detection is a process of selecting pixels similar to human skin pixels in an image, and it is widely used in fields such as face detection and gesture detection. Skin color detection is usually divided into three steps: select the color space, select the skin color model, and obtain the skin color pixel area. The most important thing is to determine the color space and skin color model.

Currently, three color spaces of RGB, YCrCb, and HSV are generally used in skin color detection.

The RGB color space uses three colors of red, green, and blue as the primary colors, and various colors are generated by combining these three colors. The RGB color space basically includes all colors that humans can feel. The RGB color space can be represented by a Cartesian coordinate system, as shown in Fig. 1. The three coordinate axes of X, Y, and Z respectively represent the three primary colors of R, G, and B. The brightness range of these three primary colors is in the closed interval of 0~1. Each point that composes this cube with a volume of 1 represents A color, such as (0,0,0) means pure black, (1,1,1) means pure white[6].

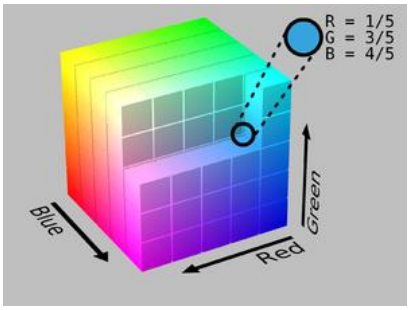


Fig. 1 RGB color space

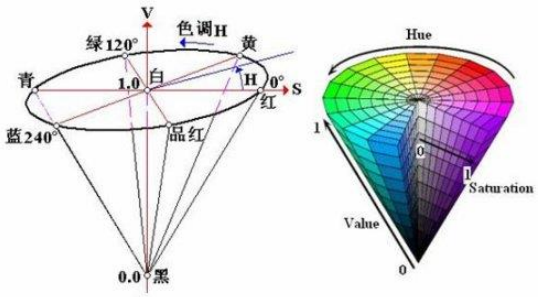


Fig. 2 HSV color space

Compared with RGB, the HSV color space is closer to human perception of color, and it keeps the calculation simple. The HSV color space is composed of three components, H, S, and V, which represent Hue, Saturation and Value respectively. Hue is the basic attribute of color, that is, the type of color. The value range is between 0 and 360°. In particular, for achromatic colors, such as black, white, and gray, their H components are all 0. Saturation represents the purity of the color, and it depends on the proportion of gray in the color. The smaller the proportion of gray, the higher the saturation. The larger the proportion of gray, the lower the saturation. Brightness indicates the brightness of the color, and the value is between 0 and 100% [6]. The schematic diagram of HSV color space is shown in Fig. 2. The conversion between HSV color space and RGB color space is shown in formula (1) [6].

ISSN: 2664-9640

$$H=\begin{cases}\dfrac{(G-B)\times\pi/3}{\max(R,G,B)-\min(R,G,B)}, R=\max(R,G,B)\\ \dfrac{(B-R)\times\pi/3}{\max(R,G,B)-\min(R,G,B)}, G=\max(R,G,B)\\ \dfrac{(R-G)\times\pi/3}{\max(R,G,B)-\min(R,G,B)}, B=\max(R,G,B)\end{cases}$$
$$S=\begin{cases}\dfrac{\max(R,G,B)-\min(R,G,B)}{\max(R,G,B)}, \max(R,G,B)\neq 0\\ 0, \max(R,G,B)=0\end{cases} \quad (1)$$
$$V=\max(R,G,B)$$

YCrCb is a color space based on the perception of the human eye and is usually used for the transmission of color video signals. In the YCrCb color space, Y represents brightness, Cr represents red chroma component, Cb represents blue chroma component, and Cr and Cb are independent of each other [8]. The conversion between YCrCb color space and RGB color space is shown in formula (2).

$$\begin{aligned}Y&=0.299R+0.587G+0.114B\\ Cr&=-0.147R-0.289G+0.436B\\ Cb&=0.615R-0.515G-0.100B\end{aligned} \quad (2)$$

In practical applications, gestures are easily affected by the external environment, such as lighting and background. The RGB color space is more sensitive to brightness, while the HSV and YCrCb color spaces are less sensitive to brightness. Therefore, consider using the HSV or YCrCb color space for skin tone recognition. In a single background, the image is relatively simple, and the skin color recognition based on the HSV color space is more accurate, in which the H component distribution is relatively concentrated, but if the background is more complex, the H and S components fluctuate significantly, and those points with smaller or larger brightness Will be ignored, and the YCrCb color space is more robust, and the interval fluctuations of the Cr component and Cb component distribution are smaller [9].

In order to reduce the influence of brightness and background, while enhancing robustness, this paper combines HSV and YCrCb color space, and uses HCrCb color space to establish a skin color model. In this paper, through an experimental analysis, the value range of H, Cr, and Cb components is obtained. The experiment selects 50 hand images of the same size, and analyzes the distribution histograms of the three components of H, Cr, and Cb of all images, through statistics , The H component is basically between 0 and 45, the Cr component is between 127 and 150, and the Cb component is between 110 and 130.

2.2. Gesture segmentation

Although the HCrCb skin color model can distinguish the hand from the background to a large extent, it is still inevitable that the background area of similar color will be segmented together with the hand area, which will interfere with the recognition of gestures, in order to reduce the interference , You need to separate the background and the hand area with similar colors.

In general, when performing gesture recognition, the human hand is closer to the camera than the background, the hand area is larger than the background area with similar colors, and the background area is very random, and the connected areas formed are often scattered. Small area. Based on the above situation, this article calculates the size of the target area, and takes the largest connected area as the hand area, and then converts the image into a binary image for further processing. After determining the hand area, you need to extract the edge of the gesture. This paper uses Canny operator to detect the edges of gestures. The Canny edge detection algorithm is based on the first derivative of the Gaussian function. It is an optimized approximation operator for the product of the signal-to-noise ratio and positioning. It is used

ISSN: 2664-9640

to calculate the gradient amplitude and phase value of the image edge, and a reasonable amplitude threshold is selected to determine the edge point. Canny operator is more versatile and effective than traditional gradient operators such as Sobel operator and Laplacian operator, and has been widely used. The algorithm flow is as follows:

(1) Smooth the image. A Gaussian filter is used to convolve the original image by row and column to obtain a smoothed image $I(x, y)$. The one-dimensional Gaussian function used to construct the filter is defined as:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{3}$$

(2) Calculate the gradient magnitude and direction. Using the finite difference of the first-order partial derivatives of the 3×3 neighborhood, the partial derivatives in the x and y directions are:

$$\begin{cases} P_x[i, j] = (I[i+1, j] - I[i, j] + I[i+1, j+1] - I[i, j+1]) / 2 \\ P_y[i, j] = (I[i, j+1] - I[i, j] + I[i+1, j+1] - I[i+1, j]) / 2 \end{cases} \tag{4}$$

The gradient amplitude is:

$$M[i, j] = \sqrt{P_x[i, j]^2 + P_y[i, j]^2} \tag{5}$$

The gradient direction is:

$$\theta[i, j] = \arctan(P_x[i, j] / P_y[i, j]) \tag{6}$$

(3) Non-maximum suppression of gradient amplitude. For each point, the neighborhood center pixel M is compared with the two pixels along the gradient direction. If the gradient magnitude of M is the smallest, it is assigned to 0. Normally, a 3×3 neighborhood is used.

(4) Double threshold method to extract edges. First, the high threshold H and the low threshold L are used to process the non-maximum value suppression gradient amplitude, and the gradient amplitude less than the threshold value is assigned to 0, thereby obtaining two edge images *H(i,j)* and *L(i,j)*, connect the edge of the image in *H(i,j)*, and when connected to the end point, search for adjacent edge points in *L(i,j)* for filling, and finally get the edge image.

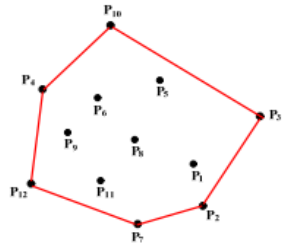
2.3. Static gesture recognition based on convex hull algorithm

This paper uses a method based on appearance features to recognize static gestures, uses a convex hull algorithm to detect the number of fingers in a gesture, and distinguishes the number represented by the gesture based on the number of fingers.

The convex hull problem can be simply described: given a plane point set $P = \{p_1, p_2, \ldots p_n\}$, find a point set S, and satisfy that the convex polygon formed by S is the smallest point set containing the plane point set P. The following respectively introduces the determination of convex polygons, the definition of convex hulls and the Graham scanning algorithm.

In a two-dimensional space, given a point set, the convex hull of this point set is (Convex Hull). The convex polygon containing the smallest area of all points in the point set is called convex hull, as shown by the red line segment in Fig. 3. A polygon is the convex hull of a point set. An actual experiment is also used to intuitively understand the convex hull model. As shown in Fig.3, imagine the points here as nails nailed on a flat surface: take a rubber band, stretch it to enclose all the nails, then loosen your hand, the rubber band will tighten on the nails, it The total length will also be minimized. At this time, the area enclosed by the rubber band is the convex hull of P.

ISSN: 2664-9640



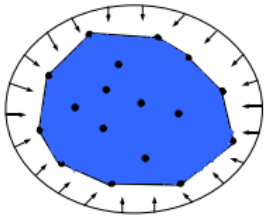


Fig. 3 Convex hull of point set P Fig. 4 Intuitive understanding of convex hull

Traditional classical two-dimensional convex hull algorithms include Jarvis March, Graham Scan, incremental algorithm, Melkman algorithm, Divide and Conquer, etc. This paper uses Graham scanning algorithm to detect the convex hull of the gesture edge image.

The basic principle of the Graham scanning algorithm is: take a certain point A in the plane, a ray AM starts from point A, and then select a length unit and the positive direction of the angle (usually counterclockwise). For any point in the plane, B can be uniquely represented by a pair of ordered numbers (ρ, θ), ρ represents the length of the line segment AB, and θ represents the angle from AM to AB. The coordinate system established in this way is called the polar coordinate system, the fixed point A is called the pole, the ray AM is called the polar axis, ρ is called the polar diameter of the point B, and θ is called the polar angle of the point B. There is an ordinal pair (ρ, θ) It is called the polar coordinate of point A.

If point A in the polar coordinate system coincides with the origin of the rectangular coordinate system, the polar axis in the polar coordinate system is the positive semi-axis of the X-axis of the rectangular coordinate system, and an angle of 90° with the X-axis counterclockwise is the positive semi-axis of the Y-axis, then the space The transformation relationship between the coordinates (ρ, θ) of any point A in the polar coordinate system and the coordinates (x, y) in the rectangular coordinate system is

$$\begin{cases} x = \rho \cos\theta \\ y = \rho \sin\theta \end{cases} \tag{7}$$

Graham scanning algorithm is a flexible convex hull algorithm, and its algorithm steps are as follows [12]:

(1) Take the point with the smallest x-coordinate. If there are multiple points with the same x-coordinate, take the point with the smallest y-coordinate, and the point must be on the convex hull.

(2) Sort the remaining points by polar angle. If multiple points have the same polar angle, the one closer to the pole is preferred.

(3) Use a stack S to store the points on the convex hull, and according to the sorting in step 2, take the first two points into the stack.

(4) Scan each point in order, and check whether the line segment formed by the first two elements on the top of the stack and this point "turns" to the right (the cross product is less than or equal to 0).

(5) If it is satisfied, pop the top element of the stack and return to step 4 to check again until it is not satisfied. Push this point onto the stack and continue to do this for the remaining points.

(6) The elements in the final stack are fixed-point sequences of convex hulls.

In order to express the features of gestures more accurately, it is also necessary to perform Convexity Defects detection on the basis of gesture contour detection. In a gesture image, the convex defect is the area between the circumscribed polygon of the convex hull of the gesture and the edge contour of the gesture. A convex defect area has four attributes, which are the starting point of convex defect detection a, the end point of convex defect detection b, and the depth point c of convex defects in the convex hull area, which is the farthest distance from the

ISSN: 2664-9640

convex hull contour point on the contour line of the gesture edge. , And the depth of convex defect |cd|, denoted by l_d. h_{box} represents the height of the circumscribed rectangle formed by the convex hull point set, and n represents the proportional relationship between l_d and h_{box} . The formula (8) is used to detect whether a convex defect is formed by a protruding finger, the formula (9) is used to calculate the distance between the starting point of the convex defect and the depth point, and the formula (10) is used to calculate the number of fingers. The key point of finger detection is the proportional relationship between the depth of the convex defect l_d and the height h_{box} of the circumscribed rectangle formed by the convex hull point set.

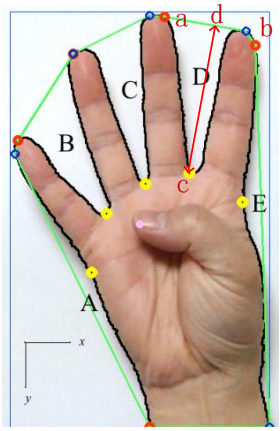


Fig. 5 Convex defect of gesture

$$count = \begin{cases} 1, s_y < b_y \text{ and } s_y < d_y \text{ and } l_d > \dfrac{h_{box}}{n} \\ 0, else \end{cases} \tag{8}$$

$$l_d = \sqrt{(s_x - d_x)^2 + (s_y - d_y)^2} \tag{9}$$

$$N = \sum count \tag{10}$$

3. Recognition results

In order to verify the effectiveness of the gesture recognition algorithm, this article conducts a gesture recognition test on the gesture images acquired by a USB camera with a resolution of 640×480. The test is divided into three parts. The first part tests the recognition rate under normal light, and the second part Test the influence of strong light and weak light on gesture recognition. The third part compares with other algorithms.

(1) Recognition rate test

Test process: Each person will test each gesture 50 times, the test gesture is 0~5, and 10 people will test separately.

Test environment: normal light.

Test results: Table 1 shows the overall average results. The average recognition rate of number 0 is 93.1%, the average recognition rate of number 1 is 93.6%, the average recognition rate of number 2 is 92.6%, and the average recognition rate of number 3 is 92.2%. , The average recognition rate of the number 4 is 91.0%, and the average recognition rate of the number 5 is 90.2%.

Table 1 Average recognition rate

Gesture category	0	1	2	3	4	5
Average recognition rate	93.1%	93.6%	92.6%	92.2%	91.0%	90.2%

(2) Light influence test

ISSN: 2664-9640

(a) Darker environment

Test process: each gesture is tested 50 times, the test gesture is 0~5, and the test results are shown in Table 2.

Test environment: dark environment.

Table 2 Gesture recognition results in a darker environment

Gesture category	0	1	2	3	4	5
Correct times	44	44	43	42	40	42
Recognition rate	88%	88%	86%	84%	80%	84%

(b) Brighter environment

Test process: each gesture is tested 50 times, the test gesture is 0~5, and the test results are shown in Table 3.

Table 3 Gesture recognition results in a brighter environment

Gesture category	0	1	2	3	4	5
Correct times	43	44	44	43	42	42
Recognition rate	86%	88%	88%	86%	84%	84%

According to the above two sets of tests, it can be seen that in a normal lighting environment, the recognition rate of static gesture recognition basically meets the needs of practical applications. Although the components that are not related to illumination are selected during gesture segmentation, excessive illumination changes will still affect the recognition rate.

(3) Algorithm comparison

In order to verify the effect of the algorithm in this paper, Cheng Xiaopeng's gesture recognition method [9] in the literature [9] is selected in the experiment. This method first extracts the hand region by using the threshold segmentation method based on the Otsu method, and then analyzes the hand Tortoise model Extract features, and finally use SVM for recognition. The average recognition rate of the comparison algorithm for gestures 0~5 is 91.7%, the average recognition rate of the algorithm in this paper is 92.1%, and the recognition rate of the algorithm in this paper is slightly higher than that of the comparison algorithm.

4. Conclusion

This paper studies the static gesture recognition based on monocular camera. In static gesture recognition, the commonly used color space is compared, and from the perspective of recognition stability, the HCrCb skin color model is selected for gesture extraction, and finally the convex hull algorithm is used for gesture recognition. Experiments have proved that the algorithm in this paper has a relatively stable recognition accuracy for static gestures in scenes with different light intensities.

Acknowledgements

Quality Engineering of Hope College, Southwest Jiaotong University(2020025).

References

[1] Gao Yaping. Research on gesture recognition method based on monocular camera[D]. Xiamen University, 2014.

[2] Liu Jun, Tian Guoguo, Li Rongkuan, Liu Xiankai. Human-computer interaction based on gesture recognition in smart space[J]. Journal of Beijing Union University (Natural Science Edition), 2010, 02:14-17+53.

ISSN: 2664-9640

[3] Qin Wenjun. Research on gesture recognition algorithm and model based on visual information [D]. Northeastern University, 2010.
[4] Zhang Huan, An Guocheng, Zhang Fengjun, Wang Hongan, Dai Guozhong. Research on human detection algorithm based on multi-color space fusion[J]. Chinese Journal of Image and Graphics, 2011, 10:1944-1950.
[5] Feng K, Yuan F. Static hand gesture recognition based on HOG characters and support vector machines[C]. Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013 2nd International Symposium on. IEEE, 2013: 936-938 .
[6] Gonzalez. Digital Image Processing [M]. Publishing House of Electronics Industry, 2003.
[7] Jiangxinyu. Conversion of color space RGB and HSV (HSL) [EB/OL]. http://blog.csdn.net /jiangxinyu/ article/details/8000999. 2012.9.20.
[8] Yu Tao. Kinect application development combat: dialogue with the machine in the most natural way[M]. Mechanical Industry Press. 2012:31.
[9] Cheng Xiaopeng. Research on gesture recognition technology based on feature extraction [D]. Wuhan University of Technology, 2012.
[10] Graham R L. An efficient algorith for determining the convex hull of a finite planar set[J]. Information processing letters, 1972, 1(4): 132-133.