ISJCRESM
https://doi.org/10.5281/zenodo.15013524
ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Jan 2024

Virtual Mouse Control via Real-Time Gesture Recognition

S. Pothalaiah, M.Purushotham, S.Dinesh, P.Santosh

Vignana Bharathi Institute of Technology, Hyderabad, India

pothalaiahs@gmail.com, purushothamreddymudupu7@gmail.com samudraladinesh11@gmail.com sunnysantosh1709@gmail.com

Abstract- Gesture recognition technology is an efficient, direct and natural human-computer interaction method, which is an active research topic in the field of computer vision. At the same time, the high degree of freedom of the hand leads to the characteristics of complex gestures, diverse changes, and high-speed movement, which makes gesture recognition technology one of the most challenging topics. In this paper, different gestures are recognized in the HSV and YCgCr colour spaces, and the detection and feature extraction based on Python are realized with the help of OPENCV2.3.1 visual library and Microsoft Visual Studio. In this paper, the related technologies of natural gesture recognition based on computer vision, such as gesture image acquisition and pre-processing, gesture image segmentation and feature extraction, and gesture modelling, etc., are proposed, and a solution suitable for virtual mouse system is proposed.

KEYWORDS: Skin Tone Detection, Gesture Recognition, Gesture Segmentation, Opencv Library, Virtual Mouse.

I. Introduction

At present, the keyboard and mouse are the interface between us and the computer, and we hope to have a more concise, intelligent and humanized human-computer interaction mode. The computer collects human video signals in front of the computer without contact, and analyses the video signals by using the relevant theories of computer vision, so as to distinguish human actions and understand human intentions, that is, we hope that the computer can "understand" us as much as possible. Gestures are a natural, intuitive, and easy-to-learn means of human-computer interaction. Gesture research is divided into gesture synthesis and gesture recognition, and gesture recognition technology has always been an important research topic of human-computer interaction. Compared with other input methods, the direct use of human hand as an input means has the characteristics of naturalness, simplicity, richness and directness. This illustrates that using computers to recognize gestures provides a more natural human-machine interface for people. However, due to its difficulty, the current research results have a low degree of practical application. After investigation, most of the domestic use of similar data gloves or through the sensing node and the convergence control node to form a system and place inertial sensor equipment such as

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Jan 2024

speedometer, gyroscope and magnetometer, etc., to track and imitate the movement trajectory of the hand, this is a set of integrated motion capture system, requires a plurality of components to work together. A small part adopts laser positioning technology, and its basic principle is to install the device of several emitting lasers in the space by calculating the angle difference, obtain the three-dimensional coordinates to complete the capture of action.

II. Protocol Analysis

In this design, we use gesture input technology based on computer vision. So, we're going to get the video stream through the camera. Then, the system detects the interactive gesture to each frame image, extracts the key frame that satisfies the condition, segments this gesture from the video signal, and then, selects the appropriate gesture feature to carry out gesture analysis, and the analysis process needs to extract the model feature parameter. Finally, gestures are classified according to model parameters and gesture descriptions are generated according to specific needs to drive related applications. Each functional module is analysed below. The design of Human-based contactless virtual mouse will be described in three aspects: gesture segmentation, gesture recognition, and mouse movements.

III. Hand Gesture Recognition Process

Gesture recognition is a process where a computer system interprets and understands human gestures, typically performed by the hands, to control or interact with the system. In the context of the AI Virtual Mouse project, the hand gesture recognition process involves several steps:

Data Acquisition: Video frames are captured from a camera feed, focusing on the region where hand gestures are expected to occur.

Preprocessing: The captured frames undergo preprocessing steps to enhance the quality of the image and reduce noise. This may involve operations such as noise removal, image resizing, and normalization.

Hand Detection and Segmentation: Techniques are applied to detect and isolate the hand region within each frame. This step is crucial for accurately recognizing hand gestures amidst varying backgrounds and potential occlusions.

Feature Extraction: Relevant features are extracted from the segmented hand region to represent different gestures effectively. These features capture distinctive aspects of hand shape, movement, texture, and spatial relationships.

Gesture Classification: A machine learning model is employed to classify the extracted features into predefined gesture categories. The model predicts the most likely gesture based on the learned patterns and classification criteria.

Post-processing: The predicted gesture may undergo post-processing steps to refine the classification results and ensure smooth interaction with the virtual mouse system. This may involve filtering noisy predictions, incorporating temporal context, or performing gesture smoothing.

3.1.1 HSV, YCgCr colour space

There are many colour spaces at present, and there are also many colour spaces applied to skin detection, such as normalized RGB colour space, YIQ, CIE-Lab, and CIE-

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Jan 2024

Luv. In the RGB colour space, the three colour components all contain luminance information, and there is a correlation, so the brightness adaptability of the algorithm is not good for skin colour detection. The HSV and YCgCr colour spaces are widely used through surveys [3]. The HSV colour space is a conical model, in which hue (H) is the basic property of colour, and the value is taken from 0~360 degrees. Saturation (S) is the purity of a colour, with higher colours being purer and lower being grey. The value in the example is 0~100%. Brightness (V) is also called brightness, and the value is 0~100. Compared with other colour spaces, the clustering characteristics of skin colour in HSV space are more obvious, and through investigation and experiments, we conclude that when the value range of H is 7~29, it is defined as the skin colour range. Although the HSV space has a high skin colour clustering characteristic, but for the background colour similar to the skin colour, the HSV space still cannot be well distinguished, so it is necessary to further carry out skin colour extraction, and the design adopts the skin colour extraction of the double colour space, mainly to improve the accuracy of skin colour confirmation, and the second space applied to is the YCgCr colour space. The YCgCr colour space, based on the YCbCr space, was proposed by De Dios. Firstly, skin colour is projected in the CG-CR colour space, the pixel area is fitted, the non-skin colour in the image is preliminarily removed, and the image of the first segmentation of skin colour is obtained, and the experiment proves that this fitting can play a good role in removing non-skin colour pixels, and the algorithm improves the speed of gesture detection, and has a higher detection rate and a lower missed detection rate. where Y still represents the brightness in the colour information, Cr still represents the difference between the red component R and the brightness Y, and Cg represents the difference between the green component G and the brightness Y. The conversion of RGB to YCgCr is as follows:

$$\begin{bmatrix} Y \\ Cg \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -81.085 & 112.000 & -30.915 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

The skin colour obtained from the survey and experiment is satisfied in the range of YCgCr space.

$$\begin{cases} Y \in [30 \ \ 235] \\ Cg \in [90 \ \ 127] \\ Cr \in [129 \ \ 173] \end{cases}$$

The experimental results showed that the skin colour extraction using the dual colour space of HSV and YCgCr could better remove the background non-skin colour area, and the comparison results of the binary value plots of the three are as follows (Fig. 2). According to the comparison chart, it can be seen that compared with the independent HSV and YCgCr space detection, dual-space skin colour detection can better remove environmental background interference, especially for the removal of skin-like backgrounds, so this design uses double-layer colour space detection [4] to provide more complete input for gesture recognition.

3.1.2 HSV, YCgCr colour space

Image processing is an important part of gesture recognition [5], and a good image processing process can provide a better recognition environment for the recognition process and reduce interference and errors. The image processing process of this design includes: corrosion, expansion, filtering, polygon approximation, etc. Filtering includes median filtering, removal of small areas, and polygon approximation [6]. The process of removing a small area uses the traversal method to query the maximum area and keep it, and delete the rest, The purpose of this operation is to remove the background noise of the face with a small area relative to the human hand (the original face is relatively far away from the camera, and the area can be smaller than the human hand surface

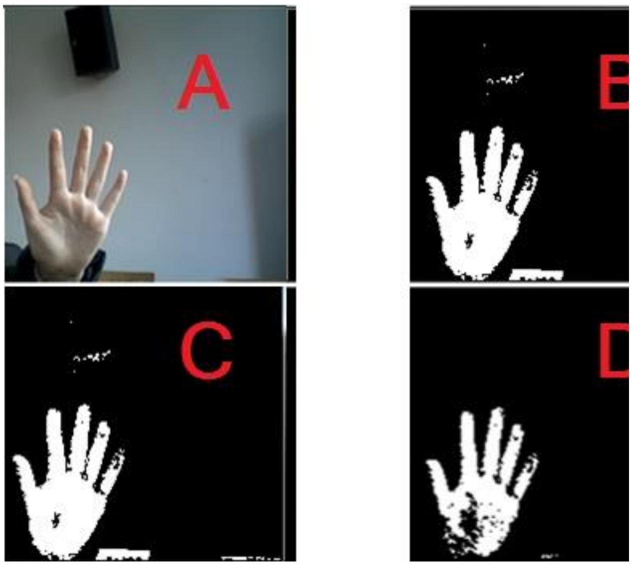

Figure 2. (A) Original; (B) HSV test results; (C) YCgCr test results; (D) Double space test results.

Figure 3. Double skin extraction results

Figure 4. Corrosion results

Figure 5. Expansion result

Figure 6. Median filter results

From the above illustration, you can clearly feel the image processing process of this design, which can basically range from the video stream to a relatively complete gesture image.

3.1 Gesture segmentation

There are many approaches to gesture recognition, which are closely related to the gesture representation (modelling) used [7]. In this paper, the centre of gravity of the gesture binary image is used to represent the palm, and the convex defect is used to detect the fingertips, and different gestures are distinguished by the different ratios of the convex contour area to the true contour area Q of different gestures. As shown in the figure, through continuous testing, the Q value of the fist can be obtained as 1~1.2, and the Q value of the five fingers should be greater than 1.2, and the Q value of single finger and two fingers can be further subdivided by testing (Fig. 9~Fig. 15).

At the same time, we apply the MeanShift algorithm [8] to track the centre of gravity and provide the possibility for the next mouse action. The idea of the algorithm can be simply understood as using the gradient climb of probability density to find the local optimum. All it does is enter a range in the image and iterate (towards the centre of gravity) until the requirements are met [9]. Through the above algorithms, we can provide the judgment conditions for the gesture judgment signal for the next mouse actions, and prepare for the next step.


Figure 7. Remove small area results


Figure 8. Polygonal approximation results

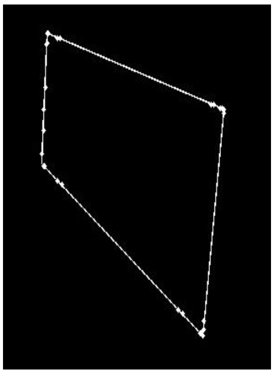
Figure 9. Fist Q value less than 1.2 (left)

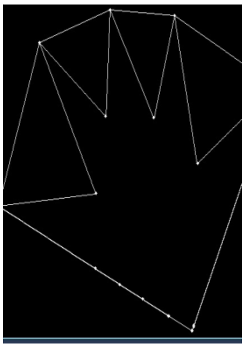
Figure 10. Five fingers spread Q value greater than 1.2 (right)

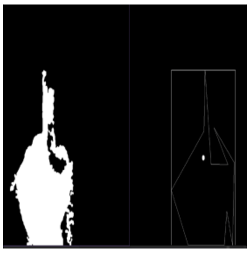
Figure 11. An extended Q value of about 1.05 (left)


Figure 12. Two refers to the extended q value of about 1.09 (left)


Figure 13. Three refers to the extended q value of about 1.13 (left)


Figure 14. Four fingers extended q value of about 1.18 (left)

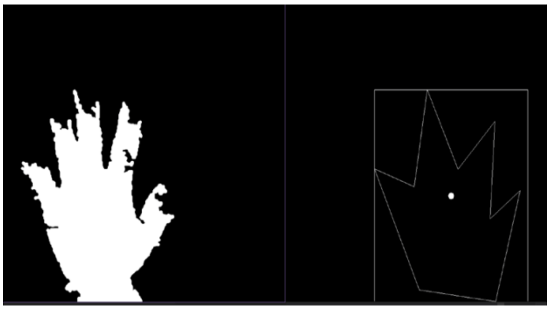
Figure 15. Five fingers extended q value of 1.25 (left)

3.2 Mouse actions

Mouse actions are the final part of this design. This design and edit two programs to realize mouse action. The mouse program realizes the left-click mouse action and the mouse tracking process to simulate the mouse function, which can be verified by the drawing tool. The virtual mouse program can be realized by combining with the mouse event (MOUSEEVENTF_LEFTUP/DOWN, 0, 0, 0, 0) and the mouse event (MOUSEEVENTF_MOVE, x, y, 0, 0), where x and y represent the relative displacement respectively, which can be obtained by the distance of the hand from the previous position on the image, and the increment of mouse movement is generated. However, because the resolution of the camera is often smaller than the resolution of the screen, the accuracy is not high. For the sake of the naturalness of the interaction, whether the mouse is pressed or not is determined by gestures, the five fingers extended means that the left button is not pressed, and the fist means that the left button is pressed. At the same time, the virtual mouse program, on the basis of the use of mouse_event (), can also obtain the tracking and imitation of gestures to achieve a comparable representation result, the specific representation results are shown in Figure 16 and Figure 17: In a large number of experimental data, some data are selected to characterize the realization stage of the results in this paper, and the following picture data have typical characteristics The results show the response time, and it is concluded that the response time of skin colour detection in gesture tracking imitation can better present stable characteristic results in a certain period of time. It is clear that the simulation of the mouse can be achieved to a certain extent with skin tone detection technology, and the data feedback is good (Fig. 18)

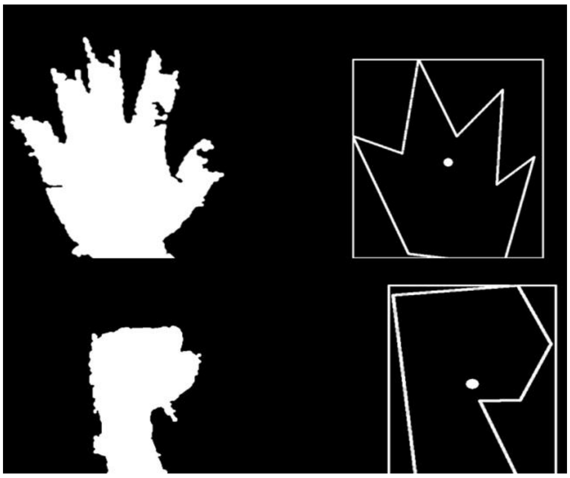

Figure 16. Five fingers open to control movement and clench fist to control left key stroke.

IV. Results and Discussions

The experimental environment of this design scheme is: Intel core i5 CPU, 2GB memory, Windows10 operating system, VS2010 + OPENCV2.3.1, and the video resolution used is 640 × 480. With the help of vs2010 software, the MeanShift algorithm is used to track the centre of gravity, and the judgment conditions of gesture judgment signals are provided for the next mouse actions, and then different gestures are recognized, and then combined with the opencv2.3.1 visual library, the recognition process of human hands is realized. The specific verification results of the experiment are as follows: virtual mouse drawing, which can be verified by drawing tools, that is, the mouse tracking and clicking functions are realized;

V. Conclusion

In this paper does not have high requirements for the accuracy of the intermediate results of gesture segmentation, so it can adapt to the changes of the environment. Through a large number of reproducible and accurate results, it is obvious that the skin tone detection technology in this paper can initially achieve a better experience in the results of tracking and imitating gestures, and it can be concluded that the proposed method has a certain degree of feasibility to be realized. This shows that the skin colour detection technology has obvious recognition of one-handed gesture recognition under monocular vision under HSV and YCgCr spatial detection, and has quite good recognition ability for translation and plane rotation, but there are still shortcomings, such as the lack of gesture samples and the limited research time limit the program optimization, so that the time operation is still limited and cannot be achieved within 24 ms. There are reasons for optimizing the program and achieving the level of interaction - the detection of convex defects requires the gesture to be kept at a certain level with the screen in order to obtain better input, but also the hardware equipment - monocular vision cameras cannot get depth information, computer memory is limited, etc. Therefore, our next step is to continue to improve the method, appropriately increase the type samples of gestures to supplement the richness and completeness of this experiment, consider adding multi-hand gesture recognition and accurately tracking and imitating gestures to obtain trajectory data, and improve the program response speed.

References

[1] Abhilash S S, Lisho Thomas, NWCC (2018) Virtual Mouse Using Hand Gesture. International Research Journal of Engineering and Technology (IRJET)

[2] Bakar MZA, Samad R, Pebrianti D, et al (2015) Finger application using K-curvature method and Kinect sensor in real-time. In: technology management and emerging technologies (ISTMET), 2015 international symposium on. Pp 218–222.

[3] Banerjee A, Ghosh A, Bharadwaj K, Saikia H (2014) Mouse control using a web camera based on colour detection. arXiv Prepr arXiv14034722.

[4] Cai Z, Han J, Liu L, Shao L (2017) RGB-D datasets using microsoft kinect or similar sensors: a survey. Multimed Tools Appl 76:4313–4355.

[5] Cao Z, Hidalgo G, Simon T, et al (2018) OpenPose: realtime multi-person 2D pose estimation using part affinity fields. arXiv Prepr arXiv181208008.

[6] Chen Z, Kim J-T, Liang J, et al (2014) Real-time hand gesture recognition using finger segmentation. Sci World J 2014:

[7] Coroiu ADCA, Coroiu A (2018) Interchangeability of Kinect and Orbbec sensors for gesture recognition. In: 2018 IEEE 14th international conference on intelligent computer communication and processing (ICCP). Pp 309–315.

[9] Ge L, Liang H, Yuan J, Thalmann D (2018) Robust 3D hand pose estimation from single depth images using multi-view CNNs. IEEE Trans Image Process 27:4422–4436.