UDC 004.62

RESEARCH ON GESTURE RECOGNITION BASED ON SUPPORT VECTOR MACHINE

Z.M. LIAO, gr.167011, master

Belarusian State University of Informatics and Radioelectronics[1], Minsk, Republic of Belarus

Tsviatkou V.– PhD in Technical Sciences

Annotation. In response to the problems of low recognition accuracy and large computation of traditional neural network algorithms, a gesture recognition detection model is designed by using human skin color features and SVM model with gesture classification recognition as the target. The method adopts bilateral filtering and other graphical processing to smooth the edges of the palm, detect and binarize the skin color of the region, and filter the binarized image to smooth the edges. The skin color space is transferred from the RCB space to the YUV space under which the gesture region is separated from the background, and morphological processing techniques are introduced in terms of gesture integrity to effectively fill the black hole region and remove the white dot region in the gesture picture, and directly edge the gesture picture. After obtaining the standardized image, the feature values are obtained. The experimental results show that the method improves the accuracy of gesture recognition compared with traditional algorithms.

Keywords. Gesture recognition, OpenCV library, Support vector machines.

Introduction

The main purpose of gesture recognition is to make it easier for users to communicate with machines, especially for people with disabilities who are physically challenged. Early scholars mainly used sensors to collect gesture data, and although this method can obtain a high accuracy rate, the hardware is generally expensive. In recent years, the technology in the field of computer vision has developed rapidly, and scholars from various fields have started to flock to the embrace of computer vision. Research can be broadly divided into two categories, one is based on artificial feature extraction gesture recognition "", this method requires human to extract gesture features, and then processing, this method is susceptible to the impact of external factors, the background color changes will also bring a large error on the results; the other is based on deep learning gesture recognition research P, using deep convolution and recurrent neural network to achieve, the method uses four deep convolutional layers to automatically perform feature learning in the original data.

System Design

The first step of the gesture recognition system is to extract the contour lines of the hand after recording the gesture from the computer camera, and then program the image pre-processing process by using the OpenCV computer vision library to improve the API to denoise the image. The denoised gesture images are subjected to skin color detection and color space conversion, as well as binarization of the images, and then further processing such as erosion and expansion is done to finally extract the overall contour of the target object. The overall flow chart is shown in Figure 1.

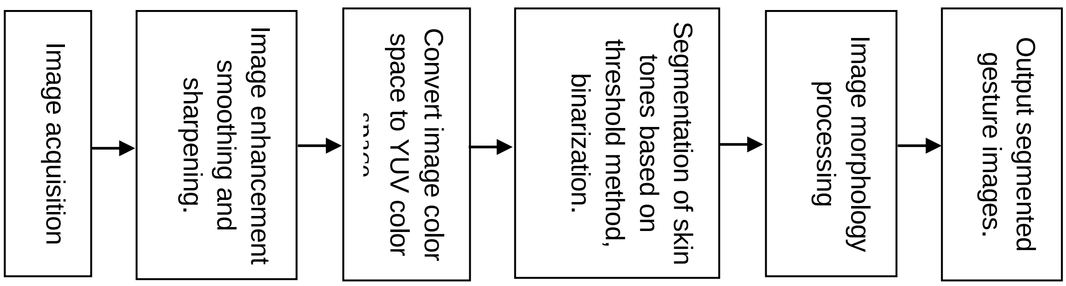

Figure 1. Flow chart of gesture segmentation

Get gestures and denoising

Due to the complex and diverse background environment of the captured images and the different performance of the capture devices, the raw gesture images captured directly from the computer camera contain a lot of noise interference, and usually need to perform pre-processing operations to maintain the integrity of the raw image information.

The OpenCV environment is configured in the PyCharm compiler, and the gesture image is recorded in the main.py source file, and the video Capture function is called to capture the camera and obtain a live gesture image from the camera. The directly acquired gesture images are affected by environmental factors, such as light intensity, gray background, etc., and need further processing to ensure that the images have valid information and key feature points, as well as to remove features from the original images [1] that have an impact on the effect or are not relevant to our target information. The noise reduction of the original image is also called filtering, and the more commonly used filtering algorithms in current research are mean filter, Gaussian filter, bilateral filter and so on. Here, the bilateral filtering algorithm is chosen to perform noise reduction on the original image.

The essence of bilateral filtering is to combine the space of pixel fields and intervals with similar pixel values to achieve the purpose of edge preservation and denoising. The closer to the center point of the coordinates of the pixel point of the gray value accounted for a greater weight, to ensure that the boundary will not be blurred out, it is a simple, non-iterative and local algorithm. Bilateral filtering method achieves the purpose of edge pixel point keeping and noise reduction by setting the corresponding function internally.

Gesture Skin Tone Detection and Binarization

To achieve the ultimate gesture recognition, one of the more important steps is gesture segmentation, which is a prerequisite for gesture recognition. Here, we use the method of skin color detection to segment the gestures. There are many gesture features [2], among which hand color is an important feature and has strong robustness, which is an inherent feature of the hand.

OpenCV uses the cvColor() function to convert to RCB space, and then uses the split() function to obtain the RGB value of each pixel in the image, which is a two-dimensional matrix split into a three-dimensional matrix. If not, the mask is set to black. In addition to the RGB color space model, there is also a more common color space is YUV, OpenCV cvtColor function parameters to CV_YUV2BGR_I420, the color space will be converted to YUV space.

SVM model-based gesture recognition

For a limited number of learning samples, by continuously reducing the structure of the problem, the learning machine can obtain the best classification ability and finally can use statistical methods to successfully predict unknown outcomes. The generalized world can find the best function in the same kind of function set, but it is difficult to find other kinds of functions. The above problem can be better solved by choosing a partial set SVM of the eigenvectors of the sample group instead of the SVM of the sample group partition [3]. The SVC classification is trained as follows.

SVC function parameters are used.

'kernel':('linear', 'rbf'), 'C': [1, 3, 5, 7, 9, 11, 13, 15, 17, 19].

'gamma':[0.00001,0.0001,0.001,0.1,1,10,100,1000] which, C is the penalty term, the larger its value, the more it will directly affect the penalty value, the training samples will also get higher accuracy; kernel indicates different kernel function types; gamma indicates kernel function coefficients, the parameters are not fixed.

Comparative analysis of experimental results

The dataset and training set are imported into the SVM model and ANN for training and testing, and the accuracy of each round and the final average accuracy are obtained as shown in Table 1 and Table 2.

Table 1- Recognition accuracy of the training and data sets for SVM

Round/Accuracy	Round1	Round2	Round3	Round4	Round5	Round6	Round7	Round8	Round9	Round 10
Training accuracy	0.987	0.984	0.991	0.987	0.997	0.987	0.987	0.987	0.987	0.987
Test accuracy	0.972	1.0	0.944	1.0	0.943	1.0	1.0	1.0	0.971	1.0

Table 2- Recognition accuracy of the training and data sets for ANN

Round/Accuracy	Round1	Round2	Round3	Round4	Round5	Round6	Round7	Round8	Round9	Round 10
Training accuracy	0.997	0.997	0.997	0.991	0.997	0.997	0.997	0.771	0.997	1.0
Test accuracy	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.771	1.0	0.971

From Table 1 and Table 2, the recognition accuracy of SVM is 0.983 and 0.988 for training and test sets. 0.974 and 0.973 for training and test sets of ANN. SVM has higher recognition progress than ANN.

Conclusion

The experimental results show that the method has improved the accuracy of gesture recognition compared with the traditional algorithm. The gesture segmentation using the skin color recognition technique is susceptible to interference in complex real-world situations, and objects with similar skin color and shapes similar to the gestures are likely to appear in complex environments, which affects the accuracy rate to a certain extent. Only nine common hand gestures were analyzed, and there are limitations in the application scope. In the future research, the model can be further improved and extended to train more gestures to improve the adaptation range.

References

1.Haralick R M, Zhuang X, Lin C, et al. The digital morphological sampling theorem[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1989, 37(12): 2067-2090.

2.Saravanan G, Yamuna G, Nandhini S. Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models[C]//2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2016: 0462-0466.

3.Yin X, Xie M. Hand gesture segmentation, recognition and application[C]//Proceedings 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation (Cat. No. 01EX515). IEEE, 2001: 438-443.