UDC 621.391

A HYBRID CLASSICATION ALGORITHM BASED ON SVM, ANN AND KNN FOR GESTURE RECOGNITION

Z.M. LIAO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 21, 2023

Abstract. A hybrid model based on SVM, ANN, KNN is proposed, which improves the accuracy compared to the traditional gesture recognition algorithm by combining it with the traditional gesture recognition algorithm.

Keywords: gesture recognition, classification algorithms.

Introduction

Human-computer interaction (HCI) technology has developed rapidly in recent years. There are many kinds of interaction methods, among which, camera-based gesture recognition is mostly used with natural interaction methods. Gesture recognition is used in a wide range of fields, including games, medical care, human-computer interaction, etc. Gesture recognition is a research area of dynamic PC, machine level learning. Gesture recognition, as a typical method of human PC communication, is an area where many experts in academia and industry are working to develop various applications to make association easier and beneficial without wearing any additional devices.

With the development of modern technology and more intelligent lifestyles, people are increasingly longing for a more natural way of interaction. Therefore, it is of great significance to research a natural and comfortable human-computer interaction mode for this phenomenon. At present, several commonly used ways of human-computer interaction are keyboard-based input, mouse-based input, voice, facial expression and gesture recognition technology. The characteristics of natural, intuitive, concise, humanized and flexible gestures can fully stimulate the potential of the hand, without the hand attached to the mouse, keyboard and other external devices, to avoid limitations. And gestures can be reflected in the process of human-computer interaction in a natural and simple way. Gestures have nothing to do with language, and their meanings are common even in different cultures and customs. In addition, there will be communication barriers in both graphic interaction and natural language interaction. Gesture[1] can play a bridging role to solve this problem well. For example, the user does not know the specific use process when using the computer keyboard and mouse, but for how to use and express gestures, it can quickly grasp. Human gestures have diversity and variability, and the efficiency of human-computer interaction can be significantly improved by endowing gestures with specific connotations and inputting them into the computer. Therefore, through the research of gesture recognition, we can better develop a natural and efficient human-computer interaction mode[2], which is in line with the current background needs.

This paper combines the advantages of SVM, ANN and KNN to propose a combination of the three classification algorithms. The small training samples of SVM can overcome the optimization of KNN K-values, ANN improves the running speed by forward propagation, and in addition ANN has a higher accuracy rate. The dataset formed by extracting the eigenvalues of the gesture images is put through a hybrid model, and the respective predictions of SVM, KNN and neural network are voted on before the result is derived.

Gesture segmentation and feature extraction

As images are disturbed and affected by conditions such as light intensity, external environment, differences in equipment performance and various noises (e.g., thermal noise, pretzel noise, etc.) during generation and transmission, the quality of the captured images can be seriously affected, thus affecting the accuracy of gesture recognition. Therefore, before the images are analyzed and processed, they must be pre-processed to improve the quality of the images, enhance the valid information in the images and reduce the useless information in the images to ensure that the results of subsequent image processing are accurate and reliable.

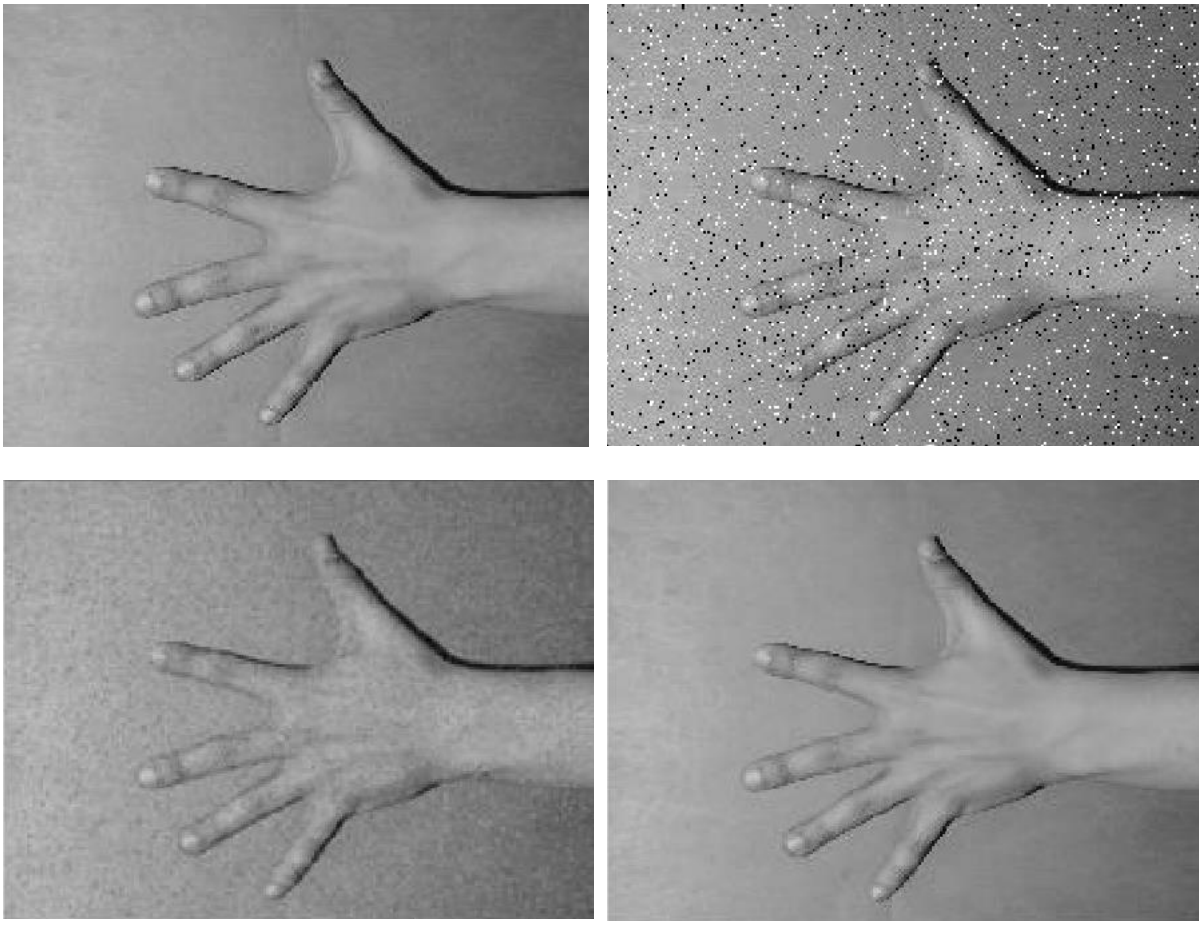

Figure 1. Comparison chart of image smoothing

Before performing gesture recognition, the gesture in the image is segmented and the feature vector of the gesture is extracted. However, due to the influence of external environment and equipment during image acquisition, the input image needs to be processed appropriately before performing gesture segmentation to improve the accuracy of the results. This chapter introduces the common methods about image pre-processing[3-4], such as median filtering and mathematical morphology processing, and gives the results after median filtering and morphology processing.

Theoretical research on traditional classification algorithms

The traditional image recognition process is to first acquire the input image, then detect and segment the desired part of the image (pre-processing), then analyze the features, and finally recognize the image. The content of image pre-processing and image segmentation [5–6] as well as feature extraction have been described earlier. The very important step in the gesture recognition process is the selection of a suitable classification model, and several different classification recognition methods will be presented next.

The KNN algorithm is a traditional machine learning method and a frequently used classical algorithm. It is statistically based, selecting the largest number of samples among the k nearest neighbors

in the experimental sample set and classifying them according to their characteristics. The basic algorithm flow is as follows:

Assume that all sample sets are defined on an N-dimensional space, and usually each sample x can be classified by the feature vector $\{a_1(x), a_2,(x)..., a_r(x)\}$ to represent, where $a_i(x)$ denotes the itch eigenvalue of sample x. Then the similarity between two samples x_i and x_j is generally calculated by using the Euclidean distance, and the formula is shown below:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2}. \quad (24)$$

The Euclidean distance formula is used to calculate the distance between two samples, and if the distance value is smaller, it means that the degree of similarity is high and the probability of belonging to the near neighbor between two samples is high, and vice versa, it means that the degree of similarity is low.

Artificial Neural Network (ANN), a mathematical tool abstracted after the brain, is generally simulated as a system consisting of many processing units (i.e., neurons) interconnected with each other. Each of these neurons contains a specific output function, called the excitation function. The connection between each two neurons represents the coefficient weighted value of the signal as it passes through that connection, called the weight. The output of the network is determined by the way the network is connected, the weights and the excitation function. Artificial neural networks are widely used in image classification and recognition systems because of their powerful learning ability. The learning and recognition process of neural networks is the process of updating and adjusting the weights of each neuron.

Support vector machines (SVM) are derived from the theory of statistical methods, which is understood to be based on the exploration of small samples. Therefore, to understand the principles of support vector machines one must be familiar with the theory of statistical methods. SVM has a very intuitive mathematical expression and a clear and explicit mathematical model, and has gained popularity among researchers as soon as it was proposed. This section will focus on the theory, mathematical model and the advantages of the algorithm present in support vector machines.

SVM-ANN-KNN based model for gesture recognition

The previous sections have pre-processed the acquired images, segmented them and extracted the corresponding feature values. The next most important part is to import the extracted feature values into different algorithm models for training and testing to obtain the average accuracy of the final gesture recognition.

Firstly, SVM is used as a classifier for experiments, and the so obtained feature values are loaded into the model, then 10-fold cross validation [7–9] is performed, and the data are divided into 10 parts, one of which is used as the test set and the remaining nine as the training set. The most important part of the SVM model is the selection of parameters, which need to be set as C (penalty factor), the kernel function and the coefficients of the kernel function. The parameter C affects the distance between the support vector and the decision plane, the larger the C, the stricter the classification, there can be no error, the smaller the C, it means that there is greater error tolerance, after many experiments to compare the C set to 350.0, the Gaussian radial basis function is a local strong kernel function, which can map a sample to a higher dimensional space, the kernel function is the most widely used one, whether large The kernel function is one of the most widely used and has better performance for both large and small samples, and it has fewer parameters compared to the polynomial kernel function [10–11], so in most cases when you do not know what kernel function to use, the Gaussian kernel function is used in preference. Finally, the dataset and training set are imported into the SVM model for training and testing, and the accuracy of each round and the final average accuracy are obtained as shown in Table 1.

Table 1. **Comparison of 10-fold cross-validation recognition accuracy of four algorithms**

Accuracy / Model	Test set（%）	Training set（%）
SVM	98,8%	98,3%
ANN	97,3%	97,4%
KNN	96,9%	98,3%
SVM-ANN-KNN	99,3%	99,2%

Conclusion

The algorithm proposed in this paper is verified. First, experiments on gesture segmentation and feature extraction are conducted, and then the extracted feature values are applied to SVM, ANN, KNN and the SVM-ANN-KNN model proposed in this paper. Finally, the four models are compared and tested to compare the gesture recognition accuracy of the four models, and the results show that the accuracy of the SVM-ANN-KNN model proposed in this paper is 99.3%, which is higher than the other three classifiers.

References

1. Karaçalı, B., Ramnath, R., Snyder, W. E. // Pattern Recognition Letters. 2004. Vol. 25. P. 63–71.
2. Kaymak, S., Helwan, A., Uzyn, D. // Procedia computer science. 2017. Vol. 120. P. 126–131.
3. Haralick, R. M., Zhuang, X., Lin, C., Lee, J. S. // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1989. Vol. 37. P. 2067–2090.
4. Lee, J. S. // Computer vision, graphics, and image processing. 1983. Vol. 24. P. 255–369.
5. Yin X, Xie M. // IEEE International Symposium on Computational Intelligence in Robotics and Automation. 2001. P. 438–443.
6. Saravanan, G., Yamuna, G., Nandhini, S. // International Conference on Communication and Signal Processing (ICCSP). 2016. PP. 0462–0466.
7. Tomasi, C., Manduchi, R. // In Sixth international conference on computer vision. 1998. PP. 839-846.
8. Shanableh, T., Assaleh, K., Al-Rousan, M. // IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 2007. Vol. 37. P. 641–650.
9. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. // In on The Move to Meaningful Internet Systems. 2003. P. 986–996.
10. Danielsson, P. E. // Computer Graphics and image processing. 1980. Vol. 14. P. 227–248.
11. Imandoust, S. B., Bolandraftar, M. // International journal of engineering research and applications. 2013. Vol. 3. P. 605–610.