Action Recognition Based on Pedestrian Detection Tracking and Local Optical Flow Trajectory Analysis

Dongbo Zhao[1, a], Hui Li[2]

[1]School of Electronic Engineering, Xi'an Aeronautical University, Xi'an, 710077, China;

[2]School of Electronic Information, Northwestern Polytechnic University, Xi'an, 710129, China.

[a]alien_dffy@163.com

Abstract

Action recognition is widely used in the fields of intelligent monitoring, social security and so on. Aiming at the problem of low detection rate and long usage time in action recognition, this paper proposes an action recognition method based on pedestrian detection, tracking and local optical flow trajectory analysis. First, integral channel pedestrian detection and Kernelized Correlation Filters(KCF) tracking was used to detect and track pedestrians, and calculate their respective local optical flow trajectories at a certain distance. Then we use Support Vector Machine (SVM) to train sample data to identify actions. The experimental results show that the method has obvious advantages over other algorithms in terms of time of use and accuracy.

Keywords

Integral channel pedestrian detection; KCF track; local optical flow trajectory.

1. Introduction

Motion recognition is widely used in intelligent surveillance, social security and other fields, and is now the focus of AI research. Many scholars have constantly put forward new algorithms in this field, some of which have been made into real products and have great social value. At present, there are three main methods of action recognition: 1) the movement process is divided into several important sub processes, and then the analysis of these sub processes; 2) important node judgment, analysis of typical action; 3) spatio-temporal analysis, analysis of the spatial and temporal features of different actions [1]. For the detection of moving objects, there are global optical flow method [2], edge direction change method (HCEO) [3] and visual word bag model [4]. These methods generally take longer time and require special equipment to accelerate, which often limits the application of the algorithm. After extracting the features, various classifiers can be used to train and predict the [5]. With the rise of depth learning, many scholars use deep learning to recognize the action of [6].

At present, the key factors restricting movement recognition are: 1) extraction of action time segment; 2) feature extraction. The method of global optical flow has many problems, such as complex computation and large computation. Many useless details are calculated; the method using key point feature extraction has significant advantages in extracting feature speed, but it cannot use the temporal and spatial context of motion, and there is a disadvantage of lower correct rate.

For this reason, this paper extracts an action recognition method based on the combination of pedestrian detection tracking and local optical flow trajectory, in order to solve the contradiction between speed and performance in action recognition.

2. Implementation of Action Recognition Algorithm

2.1 Algorithm Flow

First, the pedestrian detection of the integral channel is carried out in the video. When the pedestrians are detected, the detected pedestrians are screened, the pedestrian detection frame (too small or too small) is removed, and the KCF tracking method is used for the detected pedestrians (usually two people). When the distance of the two individuals is less than a certain value, it is considered as the

beginning of the motion capture. Then the moving position of the hand and the position of the leg are found. The sparse light flow of the two parts is calculated and the average path of the light flow is calculated. When the two person's position is greater than a certain threshold, the action is ended. The obtained trajectory of hand and leg optical flow is used as the recognition feature of movement to be normalized and used as machine learning training and prediction. Thus, through pedestrian detection, tracking, calculation of the average displacement of local light flow and starting point of positioning action, the computation can be reduced to a great extent, and the accuracy of the algorithm is also improved. The whole algorithm flow is shown in Figure 1.

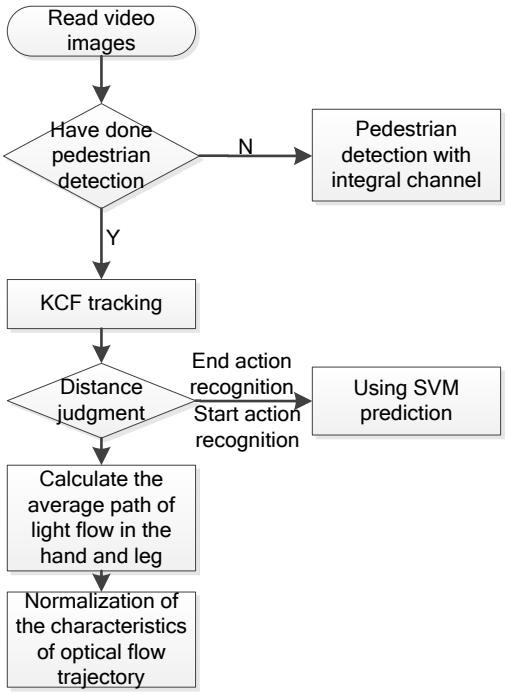

Figure 1. Whole algorithm flow chart

2.2 Pedestrians Detection and Tracking

First, determine whether the video is pedestrian detection. When detected, pedestrians track the detected pedestrians. The algorithm uses integral channel pedestrian detection [7], which can detect pedestrians very quickly and achieve real-time performance.

After detecting pedestrians, a KCF tracking algorithm with faster speed and better tracking effect is used to track pedestrians [8].

2.3 Capture Action Time and Feature Extraction

And some other algorithms have always detected the characteristics of each frame. This algorithm is extracted from two pedestrian distances from far to near to a certain range. When the action is over, the two people are far away from each other and reach a certain distance. Optical flow detection uses Pyramid LK sparse optical flow, because the selected hand and foot area is very small, and the time used for optical flow calculation is very small. After calculating the optical flow, we calculate the direction of the optical flow by calculating the mean optical flow in the direction, and the average optical flow is calculated as follows:

$$\begin{cases} x = \sum_{i=0}^{n} \frac{x_i}{n} \\ y = \sum_{i=0}^{n} \frac{y_i}{n} \end{cases} \tag{1}$$

Among them, the number of pixels in the hand and leg region represents the optical flow component in each pixel direction. After obtaining the optical flow trajectories of each region, a container is used to save the optical flow trajectories and get a sequence of time and displacement.

$$t = f(x, y) \tag{2}$$

After getting the trajectories of the 4 movements, the vectors obtained from these 4 formulas (2) form a new vector to form the frame vector of action recognition. In order to better identify the features and normalize each frame vector, the support vector machine (SVM) can be used to train the sample. The training results can be saved and the action video can be predicted.

2.4 SVM Training and Prediction

Support vector machine (SVM) is a popular machine learning method at present. This method can achieve better results for a small number of samples. After extracting the features of the action recognition video set, the SVM is trained according to various actions, and the training result is obtained. Then, the feature extraction of the test video is extracted, and the SVM is used to predict the classification of the action of the video.

3. Experiment

This experiment adopts the 64 bit operation system in win10, and develops software using VC++2015's Win32 bit and OpenCV library programming. Processor bit Intel i3-6000, the main frequency is 3.70GHz, not using multi-core parallel and GPU acceleration technology.

3.1 Processing Time

In order to achieve real-time performance of the algorithm, the computation time of each module in the statistical algorithm is needed. Pedestrian detection only detects the initial state of the video, and tracks it when the pedestrians are detected. So as long as the sum of the time used in tracking and feature extraction can achieve real-time effect, the algorithm can detect in real time. Table 1 is the time statistics used by each module of the algorithm. The experimental video uses UT-Interaction database, and the picture width is 428*312 pixels.

Table 1. Time statistics (milliseconds) for each module of the algorithm

Pedestrian detection	KCF tracking	Optical flow characteristics	SVM prediction
28	9	2	8

As can be seen from table 1, when the actual tracking and optical flow feature extraction, only about 11 milliseconds per frame, far below the real time requirement of 33 milliseconds, pedestrian detection and SVM prediction are only carried out before and after action recognition, and the use is very small.

Table 2. Time for various action recognition algorithms (milliseconds)

global optical flow	Visual word bag model	HCEO	Algorithm in this paper
69	87	13	11

Table 2 is the global optical flow method [2], visual bag model method [4], edge direction change method (HCEO) [3] and the algorithm used time contrast. As can be seen from table 2, this algorithm has obvious advantages over the global optical flow and the visual word bag model. These two algorithms need special hardware acceleration when they are applied. Compared with HCEO algorithm, this algorithm also has certain advantages.

3.2 Accuracy Detection

The same results are compared to the above three algorithms, and the reasons for the accuracy of the various algorithms are analyzed. The experiment uses The Hollywood Dataset, the video has 12 action recognition types, and each class video is 300. Select the first 50 of each class as training set, the latter 250 do the test set, and the total data of the test set is 3000. The accuracy of the statistical test set.

Table 3. accuracy rate of various action recognition algorithms (%)

global optical flow	Visual word bag model	HCEO	Algorithm in this paper
83	86	77.3	95.6

From table 3, it can be seen that the visual word bag model method has higher accuracy than the other two methods, but the extraction of SIFT features and classification by classifier makes the algorithm more used. The global optical flow method, because of the calculation of the optical flow outside the pedestrian, leads to more time and less accuracy. HCEO is less used, but the correct rate is low. The algorithm in this paper is significantly higher than other algorithms in terms of time and accuracy, and it can go to embedded devices and other real-time devices.

3.3 Algorithm Improvement Analysis

The algorithm adopts local feature extraction. After detecting pedestrians, the pedestrian is removed first, and then the area that may be active in the hand and leg is selected, as shown in Figure 2, the light flow in the area is calculated and then the path of the light flow is obtained.

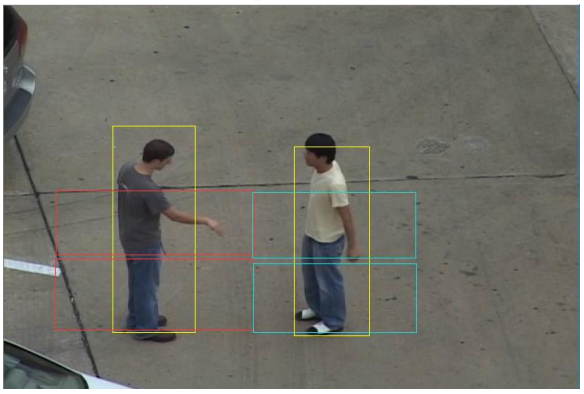

Figure 2. Area selection of light flow track in hand and leg

This method can eliminate interference from other unrelated regions and improve the accuracy of the algorithm. In the field of motion judgment, when two people close to each other and reach a certain distance, they are thought to be the beginning of the movement; when the action is completed, the two people begin to stay away from them and reach a certain distance to judge the end of the action. The path of the light flow in this period of time is calculated. Compared with other algorithms, the improved algorithm performs judgment on each frame, and judges it in a frame after leaving, which reduces most of the time.

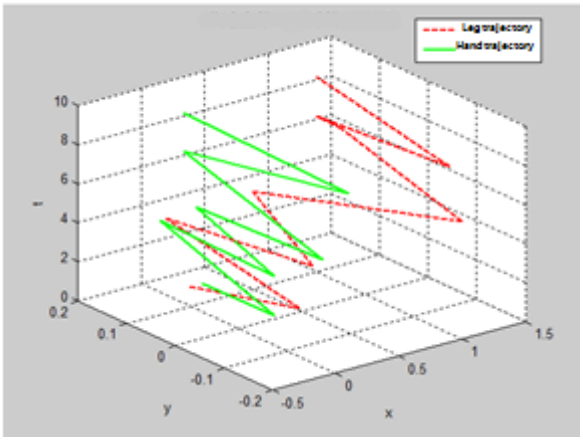

Figure 3. The relationship between the path of light flow and time

Compared with other algorithms, the method used in this paper can better distinguish all kinds of motion. Fig. 3 is the result of data simulation of kick action in a period of time. Fig. 4 is a map of the position of the optical flow trajectory in the horizontal direction.

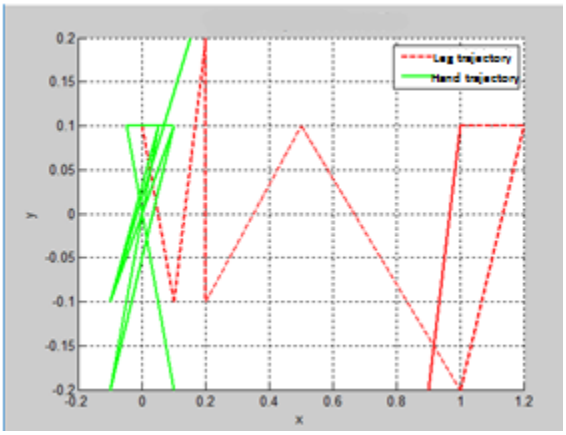

Figure 4. Plane displacement map of optical flow trajectory

From Figure 3 and Figure 4, we can see clearly that the change of leg optical flow occurs after kick kick action, and there is little displacement of the optical flow trajectory in the hand. A new feature vector is composed of a new feature vector which is calculated by the 4 regions in the picture. After SVM training, every action is judged, and it can achieve very good results.

4. Conclusion

In this paper, the method of pedestrian detection, tracking, local optical flow trajectory calculation, SVM classifier training and prediction is used to identify the action behavior. In terms of time performance and accuracy, it has obvious advantages compared with other mainstream action recognition algorithms. Through the integral channel pedestrian detection and KCF tracking, the action behavior can be described with less time and more accurate features, and the accuracy of the algorithm is improved. At the same time, the method of motion judgment is improved, and it does not need to be judged during the movement, which greatly reduces the time used in the algorithm, so that the algorithm can be transplanted on some platforms with demanding performance.

There are still some shortcomings in this paper. For example, this paper only analyzes the action of one person or two people in the video. When there are many people in the video, and the pedestrians have large images in the image, they are not analyzed. In addition, the optical flow trajectory method has a good effect on the camera fixed. When the camera is moving faster and the background noise is serious, the recognition rate will decrease. These deficiencies are necessary for future research

References

[1] Li Ruifeng, Wang Liangliang, Wang Ke review of human action recognition research [J]. pattern recognition and artificial intelligence.2014.1:27 (1).

[2] Y. Z. Hsieh and Y. L. Jeng, "Development of Home Intelligent Fall Detection IoT System based on Feedback Optical Flow Convolutional Neural Network," in IEEE Access, vol. PP, no. 99, pp. 1-1.

[3] H. M. Aye and S. M. M. Zaw, "Salient object based action recognition using histogram of Changing Edge Orientation (HCEO)," 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), London, 2017, pp. 115-122.

[4] P. Foggia, G.Percannella, A.Saggese and M. Vento, "Recognizing Human Actions by a Bag of Visual Words," 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, 2013, pp. 2910-2915.

[5] A. Cherian, B. Fernando, M. Harandi and S. Gould, "Generalized Rank Pooling for Activity Recognition," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 1581-1590.
[6] P. Foggia, A. Saggese, N. Strisciuglio and M. Vento, "Exploiting the deep learning paradigm for recognizing human actions," 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, 2014, pp. 93-98.
[7] P.Dollar, C. Wojek,B. Schiele, et al. Pedestrian detection: an evaluation of the state of the art [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 743-761.
[8] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 583-596, March 1 2015.