Scientific Research Publishing

Journal of Computer and Communications, 2021, 9, 10-18
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Study on Local Optical Flow Method Based on YOLOv3 in Human Behavior Recognition

Hao Zheng[1], Jianfang Liu[1], Mengyi Liao[1,2]

[1]Henan Intelligent Traffic Safety Engineering Technology Research Center, Pingdingshan University, Pingdingshan, China
[2]National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China
Email: liu_jianfang@126.com

How to cite this paper: Zheng, H., Liu, J.F. and Liao, M.Y. (2021) Study on Local Optical Flow Method Based on YOLOv3 in Human Behavior Recognition. *Journal of Computer and Communications*, **9**, 10-18.
https://doi.org/10.4236/jcc.2021.91002

Received: October 21, 2020
Accepted: January 5, 2021
Published: January 8, 2021

Abstract

In the process of human behavior recognition, the traditional dense optical flow method has too many pixels and too much overhead, which limits the running speed. This paper proposed a method combing YOLOv3 (You Only Look Once v3) and local optical flow method. Based on the dense optical flow method, the optical flow modulus of the area where the human target is detected is calculated to reduce the amount of computation and save the cost in terms of time. And then, a threshold value is set to complete the human behavior identification. Through design algorithm, experimental verification and other steps, the walking, running and falling state of human body in real life indoor sports video was identified. Experimental results show that this algorithm is more advantageous for jogging behavior recognition.

Keywords

YOLOv3, Local Optical Flow Method, Human Behavior Recognition

1. Introduction

As a sub-field of artificial intelligence technology, computer vision and deep learning have developed rapidly in recent years. With the variety of video scenes and the increasing variety of behaviors, the accuracy and real-time requirements of the algorithm are becoming higher and higher. Human behavior recognition technology based on video sequence has also developed from the earliest traditional classification method based on manual design features to the current method based on deep learning automatic feature extraction. The former requires manual design of features and then classification by classifier. The latter is to learn features autonomously through the neural network, then establish the se-

mantic representation of the target hierarchically, and mine the association between the data, and finally use the full connection classifier for classification [1].

When the light source is illuminated on the surface of the object, the gray level of the surface will be spatially distributed. When people observe moving objects, continuous changing images will be formed on the retina, which is called optical flow [2] [3]. The optical flow is the instantaneous velocity of the pixel motion of the spatial moving object on the observation imaging, which represents the direction and modulus of the local motion. The temporal variation and correlation of pixel intensity data in videos can be used to determine the motion of their respective pixel positions [4], so the motion can be reflected through the change of optical flow. When the human body moves, the optical flow modulus of multiple positions in the image is relatively large. The direction of the optical flow changes from the original stable to the later divergent, and the changes of the optical flow direction and modulus are both larger than the normal state.

In recent years, algorithms in image processing field have been constantly updated and optimized, and various optical flow calculation methods have been proposed. The sparse optical flow method in the traditional optical flow method only calculates individual points or points with special significance in image recognition, while the dense optical flow method proposed by Gunnar Farneback [5] calculates the dense optical flow field of the entire image by calculating the motion translation model of each pixel on the image. This method can obtain the feature of video image at pixel level through dense optical flow field, so the effect of human motion recognition based on it is obviously better than that based on sparse optical flow. However, the speed of the dense optical flow method is limited due to the large amount of pixel points and overhead. With the wide application and good effect of optical flow method in behavior recognition, the research of optical flow method in the field of abnormal video detection is getting deeper and deeper. Mehran introduced a new method that uses the Social Force model to detect and locate abnormal behaviors in crowd videos [6], which uses the spatio-temporal average value of optical flow to advection it. Simonyan *et al.* proposed a dual-stream ConvNet architecture with space and time networks, proving that ConvNet trained in the case of multi-frame dense optical flow could achieve very good performance with limited training data [7]. By replacing the optical flow with motion vector, Zhang *et al.* accelerated this ConvNet architecture, and transferred the knowledge learned from the optical flow and CNN to the motion vector CNN, which could significantly improve the performance of the latter [8]. Sun introduced a motion representation for video motion recognition, called optical flow guidance feature (OFF), which can quickly and steadily extract time information. By directly calculating the pixel-by-pixel space-time gradient of the depth feature map, OFF can be embedded into any existing CNN-based video motion recognition framework [9]. Wang *et al.* introduced a new continuous optical flow framework to capture pixel dynamics by representing a group of continuous RGB frames sequentially through

a single dynamic image, and believed that this representation could capture actions more effectively than RGB frames [10]. Leyva *et al.* proposed an online framework for video anomaly detection using the features of prospective occupancy and optical flow, with a set of highly descriptive features extracted from a novel unit structure as key points, which helped to define the anomaly area in a rough to fine way [11]. After analyzing and studying the current situation of anomaly detection in surveillance video, this paper fully recognizes the advantages of deep learning framework in anomaly detection. In this paper, YOLOv3 is combined with local optical flow method. Based on the dense optical flow method, the optical flow modulus of human target detection area is calculated, and the threshold value is set to realize human behavior recognition, so as to reduce computation and save time and expense.

2. Local Optical Flow Method Based on YOLOv3

2.1. Algorithm Design

The local optical flow method based on YOLOv3 is to localize the dense optical flow method on the basis of YOLOv3 algorithm to realize the purpose of saving running time and speeding up running speed. Human behavior recognition is realized by setting the threshold with the change of optical flow mode [12]. The process of human behavior recognition algorithm based on local optical flow method based on YOLOv3 is shown in **Figure 1**.

The specific steps are as follows:

Step 1 Video preprocessing adopts the frame-by-frame extraction method to ensure the relationship between the frames remains unchanged, and at the same time reduces the operation by setting the time interval to reduce the number of extracted image frames. Image processing uses grayscale and mean filtering to remove the data redundancy and enhance the image.

Step 2 Target detection uses YOLOv3 algorithm to obtain the coordinates and width and height of the target to calculate the target area of the image where the human body is located.

Step 3 Combined with YOLOv3 and KCF (Kernel Correlation Filter), the target tracking function of local optical flow human behavior recognition algorithm is completed by using its high tracking accuracy and robustness.

Step 4 Then the local optical fluidization is realized by dense optical flow method to calculate the optical flow modulus of the target region. This can reduce the overhead of the whole image calculation.

Step 5 Behavior descriptors are constructed to describe human behavior recognition.

Step 6 The behavior descriptors are calculated and the image threshold is set for local maximization to realize the recognition of human behavior.

- Optical flow computation

The optical flow constraint equation is the general method of optical flow

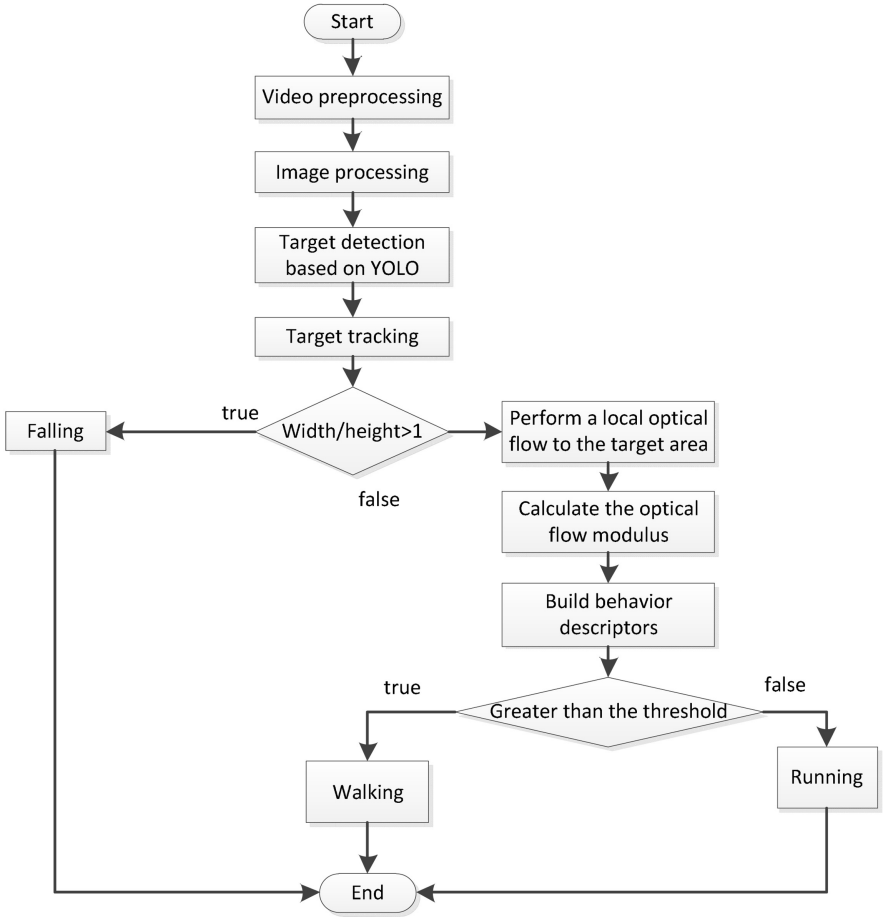


Figure 1. Flow of local optical flow method for human behavior recognition algorithm based on YOLOv3.

calculation, as shown in Formula (1).

$$\Delta I \cdot V_p + I_t = 0 \tag{1}$$

where, $\Delta I = \left(I_x, I_y\right)$ represents the gradient at the image midpoint *P*, $V_p = \left(u, v\right)$ represents the optical flow of pixel *P*, and It represents the gray change rate of the images before and after the two frames.

The key of optical flow calculation is to get the gradient of pixel point. At present, there are many methods to calculate optical flow gradient, such as Horn operator, Prewitt operator, Sobel operator and Barron operator. Among them, the classic one is the Horn operator, which calculates the optical flow gradient of pixel points by solving the first-order difference of gray level between adjacent pixel points. The key to calculation is to find the gradient of each pixel point in the image [13] [14].

- Behavior descriptors

For the input video, the optical flow of each pixel on all subsequent image frames except the first frame is calculated according to the optical flow calculation method described in (1). If the optical flow at the pixel point (*x*, *y*) on the image of the *M*th frame is expressed as ($u_{x,y,m}$, $v_{x,y,m}$), the optical flow modulus

value is calculated as shown in Formula (2):

$$M = \sqrt{\left(u_{x,y,m}^2 + v_{x,y,m}^2\right)} \tag{2}$$

The variation of optical flow modulus at pixel point (x, y) on the image of the Mth frame is shown in Formula (3):

$$\Delta M = \left|M_m - M_{m-1}\right| \tag{3}$$

- Threshold setting

In the calculation of behavior descriptors, the optical flow modulus of image pixels and the change of the optical flow modulus between the two frames are calculated by the square formula. In the local optical flow method human behavior recognition algorithm based on YOLOv3, the change of the optical flow mode value of the image in the behavior descriptor is utilized to obtain the maximum optical flow mode value of the region by local maximization. Then the average value of the maximum optical flow modulus in the video sequence image is set as the threshold value. When the optical flow mode value is greater than the threshold value, the human body is in a running state; when the motion speed is less than the threshold value, the human body is regarded as walking state. The value of W/h is calculated to determine whether the behavior is abnormal. When the value is greater than 1, the human body is in a state of falling, otherwise it is in a standing state.

2.2. Experimental Design and Result Analysis

The videos used in this chapter in the experiment are all real indoor videos of human movement in real life, including walking, running and falling.

2.2.1. Experimental Environment and Parameter

The experimental environment is shown in **Table 1**.

The training parameters: In YOLOv3 algorithm training, the image was clipped to 416 × 416, the target object threshold was set to 0.6, and the target border threshold was set to 0.8. The average value of the local maximum optical flow modulus is set to the threshold value.

Table 1. The experimental environment.

The graphics card	GeForce GTX 950M, NIVDIA
Deep learning framework	tensorflow
Image processing library	keras
Parallel computing platform	Opencv
GPU acceleration library	Cuda10.0
Development of language	Cudnnv8.5
software development environment	Python
Operation system	Windows

2.2.2. Data Preprocessing

- Frame by frame extraction

When watching videos in daily life, there is a corresponding relationship between video playback speed and image frames. When a video is shown at 15 frames per second, people can obviously feel the lag, and when it reaches 25 frames per second, people can feel the normal playing speed. Doubling speed in the software is to take advantage of the increase of playing frame speed per second, giving people the sense of double speed. The video data in this paper is played at 30 frames per second, with a total frame number of 708 and a duration of 23 seconds. In this paper, frame-by-frame method is used to extract video frames.

Frame by frame extraction method is to extract frames according to the set interval between frames to ensure that the target behavior changes and optical flow modulus values will not cause errors due to the frame-taking algorithm within a certain time interval.

- Image processing

In this chapter, image redundancy is reduced by graying and image enhancement is achieved by mean filtering.

Mean filtering is a typical linear filtering algorithm. A template is given to the target pixel on the image, which includes the adjacent pixels around it (a filter template is formed by taking the target pixel as the center and adding 8 pixels around it, that is, including the target pixel itself). Then, the average value of all pixels in the template is used to replace the original pixel value, smoothing the pixel, highlighting the details and ignoring the edges, so as to achieve the purpose of denoising.

In this chapter, video frames are processed by graying operation and then mean filtering. In this way, the noise is removed from the human body target and the target frame during the next step of target detection, so as to improve the target recognition rate and classification confidence. Image frame mean filtering processing is shown in Figure 2.

2.2.3. Experimental Results

In this chapter, the frame-by-frame extraction method is adopted to extract the previous frame and the next frame of the image according to the time frame interval, so as to ensure that the range of optical flow mode value in the optical flow operation will not cause errors due to the extraction method.

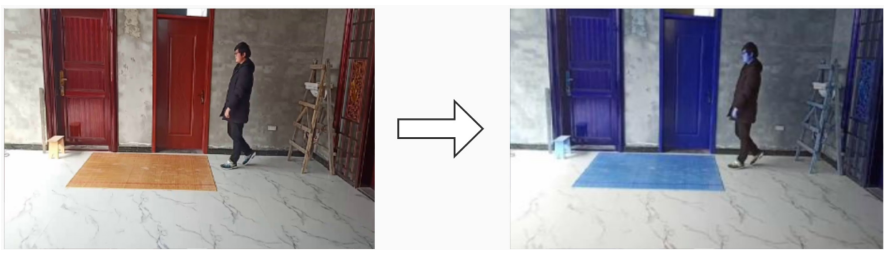


Figure 2. Mean filtering processing.

The local optical flow method based on YOLOv3 in human behavior recognition maximizes the optical flow modulus locally according to the behavior descriptors, averages the maximum optical flow modulus of each video frame, and finally sets the mean value as threshold to realize human behavior recognition.

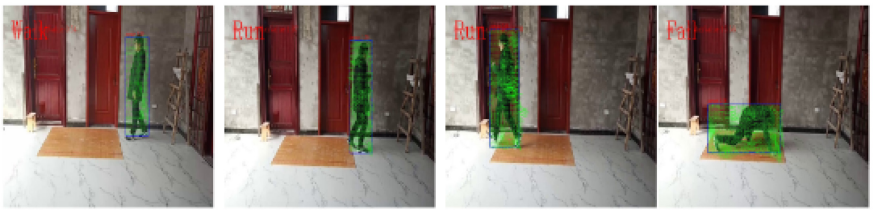
The total number of experimental video frames is 708. Video is extracted by frame - by - frame method. Local optical flow method based on YOLOv3 was used to identify frame 6, 346, 447, and 655, and the comparison of recognition effects was shown in Figure 3.

Figures 3(a)-(d) show the recognition effect of using the above algorithm: walking state, running state, running state and falling state. Since the local optical flow method has a low recognition rate due to the interference of external factors, the detection of all frames in the video recognition is as follows: the number of frames in walking state is 320 frames in total, and 229 frames are correctly identified; running state is 240 frames in total, and 185 frames are correctly identified; falling state is 148 frames in total, and 135 frames are correctly identified. The average accuracy of algorithm is 80.96%, the average precision is 91.84% and the average recall rate is 91.28%, as shown in Table 2.

3. Conclusions

The paper compares dense optical flow method and sparse optical flow method and proposes a local optical flow behavior recognition algorithm based on YOLOV3. The local optical flow method based on YOLOv3 is to localize the dense optical flow method on the basis of YOLOv3 algorithm to realize the purpose of saving running time and speeding up running speed. Human behavior recognition is realized by setting the threshold with the change of optical flow mode.

The experimental results show that the recognition rate of local optical flow method is low due to the interference of external factors. The algorithm has



(a) (b) (c) (d)

Figure 3. Identification effect.

Table 2. Evaluation indexes of local optical flow behavior recognition algorithm.

Class	Accuracy	Precision	Recall
Walk	71.56%	89.08%	91.07%
Run	77.50%	94.59%	93.58%
Fall	91.21%	91.85%	89.21%
average	80.96%	91.84%	91.28%

more advantages for jogging behavior recognition, with an average accuracy of 80.96%, an average precision of 91.84%, and an average recall rate of 91.28%.

During the experiment, it is found that the optical flow mode is increased due to the camera's slight shaking during the video recording process, which leads to the decrease of the recognition rate. The following research will focus on how to reduce the influence of external factors on the recognition rate.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Funding

This research is funded by The Science and Technology Development Program of Henan Provincial Science and Technology Department, grand number 182102310040; this research is funded by Pingdingshan University Youth Scientific Research Fund Project, grand number PXY-QNJJ-2018005; this research is funded by Henan Intelligent Traffic Safety Engineering Technology Research Center.

References

[1] Liu, X. (2019) Research and Application of Human Behavior Recognition Technology Based on Deep Learning. Beijing University of Posts and Telecommunications, Beijing.

[2] Zhang, Z.M. (2019) Research on Human Behavior Recognition Technology Based on Deep Learning. Shandong University, Jinan.

[3] Wen, M.L., Zhao, X. and Cai, M.Q. (2017) End-to-End CapTCHA Recognition Based on Deep Learning. *Wireless Internet Technology*, **14**, 85-86.

[4] Ren, S., He, K., Girshick, R., *et al.* (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[5] Mehran, R., Oyama, A. and Shah, M. (2009) Abnormal Crowd Behavior Detection Using Social Force Model. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 935-942. https://doi.org/10.1109/CVPR.2009.5206641

[6] Simonyan, K. and Zisserman, A. (2014) Two-Stream Convolutional Networks for Action Recognition in Videos. *Proceedings of the* 27*th International Conference on Neural Information Processing Systems*, **1**, 568-576.

[7] Zhang, B., Wang, L., Wang, Z., *et al.* (2016) Real-Time Action Recognition with Enhanced Motion Vector CNNs. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2718-2726. https://doi.org/10.1109/CVPR.2016.297

[8] Sun, S., Kuang, Z., Sheng, L., *et al.* (2018) Optical Flow Guided Feature: A Fast and Robust Motion Representation for Video Action Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 1390-1399. https://doi.org/10.1109/CVPR.2018.00151

[9] Wang, J., Cherian, A. and Porikli, F. (2017) Ordered Pooling of Optical Flow Se-

quences for Action Recognition. *IEEE Winter Conference on Applications of Computer Vision*, Santa Rosa, CA, 168-176. https://doi.org/10.1109/WACV.2017.26

[10] Leyva, R., Sanchez, V. and Li, C.T. (2017) Video Anomaly Detection with Compact Feature Sets for Online Performance. *IEEE Transactions on Image Processing*, **26**, 3463-3478. https://doi.org/10.1109/TIP.2017.2695105

[11] Wang, Z.L., Huang, M. and Zhu, Q.B. (2018) Optical Flow Detection of Moving Objects Based on Deep Convolution Neural Network. *Optoelectronic Engineering*, **45**, 1-9.

[12] Lin, J. and Lin, L. (2016) A Method for the Detection of Crowd Disturbance Based on the Frequency of Optical Flow Modulus Change. *Computer Science*, **43**, 283-287.

[13] Wu, Q.T., Zhou, Y.M., Wu, X.Y., *et al.* (2020) Real-Time Running Detection System for UAV Imagery Based on Optical Flow and Deep Convolutional Networks. *IET Intelligent Transport Systems*, **14**, 278-287. https://doi.org/10.1049/iet-its.2019.0455

[14] Shi, L.W., Deng, X., Wang, J. and Chen, Q.S. (2017) Multi-Target Tracking Based on Optical Flow Method and Kalman Filter. *Computer Applications*, **37**, 131-136.