Human Cross-Border Alarm Detection Method Based on OpenPose

Hang Yu[1], Qinjun Zhao[1](✉), Yong Zhang[1], and Shengjun Shi[2]

[1] University of Jinan, Jinan 250022, China
[2] Harbin Institute of Technology Robotics Group Co., Ltd, Harbin 150000, China
cse_zhaoqj@ujn.edu.cn

Abstract. Cross border detection is often used to monitor the behavior of people in specific places where people often entry and exit, people may cross some unsafe or forbidden borders, thus causing dangerous behaviors, such as large power plant or electrical equipment room. In order to prevent the occurrence of dangerous behaviors, this paper proposed a method for human cross-border alarm detection. First, the camera captures the image of the scene, and design the unsafe bounding line. Second, detect the human and its foot based on OpenPose. Third, when a cross-border behavior occurs, judge whether there is an intersection between the boundary line and the line formed by human feet in two images to send an alarm signal. This method effectively saves costs, replaces artificial ways and improves detection efficiency at the same time, and can make an alarm in time when humans cross an unsafe boundary.

Keywords: Cross-border · Detection · OpenPose · Kalman filter

1 Introduce

Human cross-border detection is often used in places with high insecurity factors, such as large power plant, electrical equipment room, etc. There may be multiple entrances and exits that can be crossed in these places, some of which are dangerous and prohibited. In order to prevent the occurrence of some dangerous behaviors, managers need to monitor people's behaviors through the monitoring system for a long time, and promptly alarm the occurrence of unsafe behaviors. However, through this method of human supervision, not only the cost of supervision is high, but also it is not possible to control all the information in real time, so the efficiency is low [1].

With the development of computer vision technology, it has become a new and effective method to analyze surveillance video by computer, capture human in the video and make judgments. The basic methods of domestic and foreign scholars on human cross-border detection include human detection and cross-border judgment. In the human detection part, [2] proposes to separate humans and backgrounds through Gaussian background modeling methods to detect humans in motion, the deep convolutional network model that has developed rapidly in recent years also has a very good effect on human

W. Fu et al. (Eds.): ICMTEL 2021, LNICST 388, pp. 469–475, 2021.
https://doi.org/10.1007/978-3-030-82565-2_39

detection [3]. The cross-border judgment part mainly analyzes the geometric positions of humans and the boundary line to judge whether humans have cross-border behaviors [4]. In order to realize the simplicity and efficiency of human's cross-border detection application, this paper combines OpenPose and Kalman filter to propose a method that can detect humans in video surveillance in real time and react to the cross-border behavior.

2 Detection and Judgment

2.1 OpenPose

Compared with the method of separating people and background through image processing and modeling to realize the recognition of moving humans, this paper uses the OpenPose model proposed by [5] to realize a more effective and fast human detection method. OpenPose is an open source library which present a real-time approach to detect the 2D pose of multiple people in an image. By using OpenPose, not only the position of the human in the image can be obtained, but also the key points of the human body can be detected. In the part of cross-border detection, it's clear that judging the geometric position of the human foot and the boundary line is more accurate and reliable than judging the geometric position of the entire pedestrian and the boundary line.

OpenPose first initializes and fine-tunes the input image through the first 10 layers of the VGG-19 network to generate a feature map F. The feature map F is calculated through two CNN network branches. Branch 1 is used to predict the confidence of the key points of the human body in the image, and branch 2 predicts the PAFs (Part Affinity Fields) of the key points. The architecture is shown in Fig. 1.

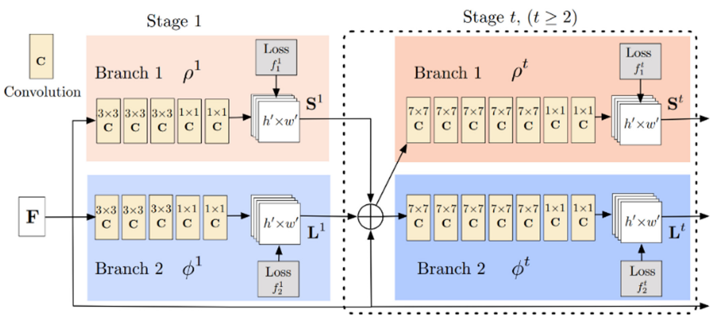


Fig. 1. Architecture of the two-branch CNN

The calculation equation of the confidence maps S^t and the PAFs L^t is as follows:

$$S^t = \rho^t(F, S^{t-1}, L^{t-1}) \quad \forall t \geq 2 \tag{1}$$

$$L^t = \varphi^t(F, S^{t-1}, L^{t-1}) \quad \forall t \geq 2 \tag{2}$$

When using formula (1) and (2) to predict key points, ρ^t and φ^t are the CNNs for inference as Stage t. Each iteration of the algorithm takes the S and T generated in the previous iteration and the original feature map F as input. In order to guide the network to iteratively predict confidence maps of body parts in the first branch and PAFs in the second branch, there are two loss functions at the end of each stage:

$$f_S^t = \sum_{j=1}^{J} \sum_{P} W(P) \bullet ||S_j^t(P) - S_j^*(P)||_2^2 \tag{3}$$

$$f_L^t = \sum_{c=1}^{C} \sum_{P} W(P) \bullet ||L_c^t(P) - L_c^*(P)||_2^2 \tag{4}$$

where S_j^* is the groudtruth part confidence map, L_c^* is the groudtruth part affinity vector field, W is a binary mask which is used to avoid penalizing the true positive predictions.

After multiple iterations, the Hungarian algorithm is finally used to integrate and connect the key points and output the skeleton structure of the human body. As shown in Fig. 2.

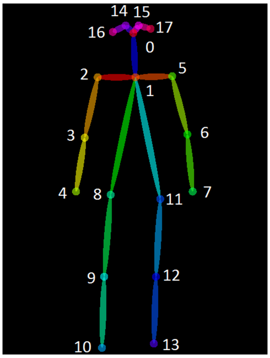


Fig. 2. OpenPose detection

2.2 Judgment

After the 18 key points of the human body are detected in the previous section, the human body needs to be tracked to achieve cross-border tracking. The Kalman filter algorithm is used here. First, draw a rectangular bounding box based on the detected key points of the human body. The bounding box contains the position, height and width of the human body. Second, send the information to the Kalman filter, and the person ID is assigned and traced.

Kalman filter use the bounding box information in the current frame to predict the bounding box information in the next frame, and the prediction can be modified according to the actual boundary box information to achieve the optimal effect [6]. The algorithm is as follows:

Predict:

$$\mathrm{x}^{\mathrm{k}} = \mathrm{Ax}_{\mathrm{k}-1} + \mathrm{Bu}_{\mathrm{k}-1} \tag{5}$$

$$\mathrm{P}_{\mathrm{k}} = \mathrm{AP}_{\mathrm{k}-1}\mathrm{A}^{\mathrm{T}} + \mathrm{Q} \tag{6}$$

Update:

$$\mathrm{x}^{\mathrm{k}} = \mathrm{P}_{\mathrm{k}}\mathrm{H}^{\mathrm{T}}(\mathrm{HP}_{\mathrm{k}}\mathrm{H}^{\mathrm{T}} + \mathrm{R})^{-1} \tag{7}$$

$$\mathrm{x}_{\mathrm{k}} = \mathrm{x}_{\mathrm{k}} + \mathrm{K}_{\mathrm{k}}(\mathrm{z}_{\mathrm{k}} - \mathrm{Hx}_{\mathrm{k}}) \tag{8}$$

$$\mathrm{P}_{\mathrm{k}} = (\mathrm{I} - K_k\mathrm{H})P_k \tag{9}$$

where x_{k} is the state of the target at k time, u_{k} is the external input, P_{k} is the error matrix, Q and R are the noise covariance matrices, H is the observation matrix, K_{k} is the Kalman gain coefficient.

When the different foot points of the target in two images have been caught, connect the foot points using straight lines. Suppose the first point is (x_1, y_1), the second point is (x_2, y_2), use (x_3, y_3) and (x_4, y_4) to represent the ends of the boundary lines. If the result d in the following equation is not zero, then calculate the results of d1, d2, d3, d4, if d1 and d2 are opposite, d3 and d4 are opposite, so there is an intersection between the two lines.

$$d = (x_2 - x_1) * (y_4 - y_3) - (y_2 - y_1) * (x_4 - x_3) \tag{10}$$

$$d1 = (x_4 - x_1) * (y_2 - y_1) - (y_4 - y_1) * (x_2 - x_1) \tag{11}$$

$$d2 = (x_3 - x_1) * (y_2 - y_1) - (y_3 - y_1) * (x_2 - x_1) \tag{12}$$

$$d3 = (x_1 - x_3) * (y_4 - y_3) - (y_1 - y_3) * (x_4 - x_3) \tag{13}$$

$$d4 = (x_2 - x_3) * (y_4 - y_3) - (y_2 - y_3) * (x_4 - x_3) \tag{14}$$

If a person crosses the boundary line, the two lines will have an intersection, then the system can detect that a person is crossing the boundary. As shown in Fig. 3.

2.3 Overall Framework

To summarize, achieve human crossing detection, OpenPose model is used to detect the person's key points in the image at first, and draw the bounding box according to the key points. Then Kalman filter is used to track and assign different human IDs based on the bounding box. Finally, the system judges that someone has crossed the boundary based on the intersection between the line generated by human movement and the boundary line and reacts. The overall framework is shown in Table 1.

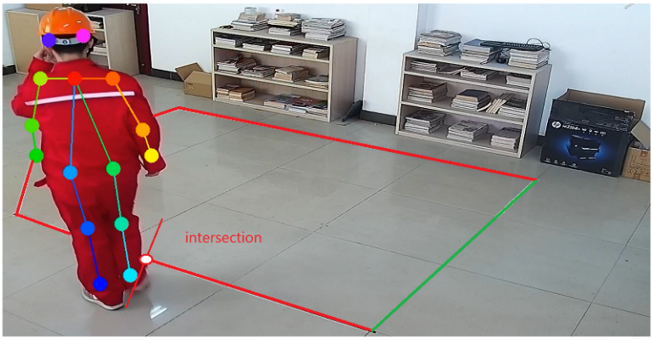


Fig. 3. Intersection occurs when human crossing border

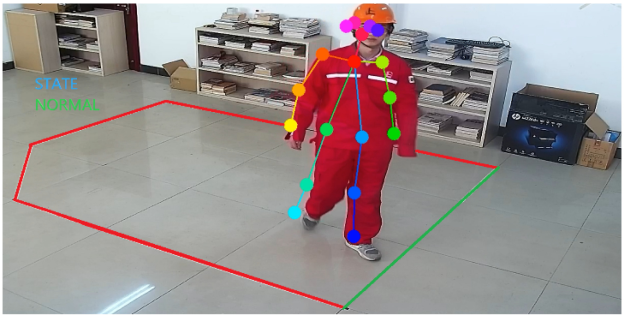
Table 1. Overall framework.

```
Load Model: OpenPose
Input: image from video stream
Repeat:
    detect key points of human's body with OpenPose
    draw bounding box of human
    track human body with Kalman filter
    if (human is detected in the image)
        record human ID and location
    then
        compare current information with the next image
        if (human illegally crossing borders)
            system alarm
    end
Until: end of video
```

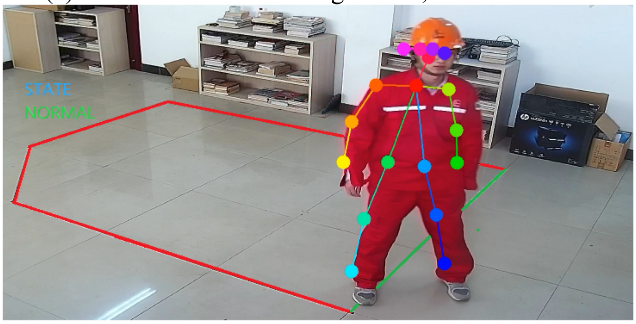
3 Experiments and Results

An irregular boundary box is designed as the test object in the experiment, the bounding box can be regular or not irregular, in fact, it depends on your actual demand. Some sides are not allowed to pass, and we mark them with red lines. Besides, one side which we mark it with green line is allowed through. In order to verify the effectiveness of the method, a variety of human behaviors are tested in the experiment, as shown in Fig. 4.

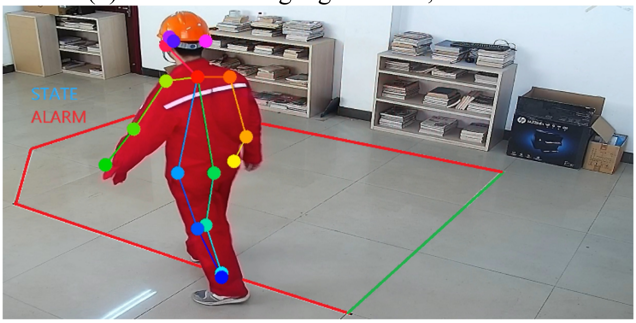
In the experiment, the weight of the OpenPose model used was provided in [5]. The experiment was run on the Intel (R) Core (TM) 9400H CPU, Ubuntu 18.04 operating system. We tested dozens of behaviors and compared their effects on a variety of models, by combining the use of Kalman filter, the final experiment achieved good detection results, and the experimental results are shown in Table 2.



(a)Human dose not crossing border, state is normal



(b)Human crossing legal border, state is normal



(c)Human crossing illegal border, state is alarm

Fig. 4. Three kinds of human state in testing

Table 2. Test result

Method	Detect accuracy
G-RMI[7]+KF	89%
SSD[8]+KF	91%
Ours (OpenPose+KF)	95%

4 Conclusion

In this paper, the human cross-border detection method based on OpenPose and Kalman filter is proposed. Compared with other methods, this method is simple and does not require complex image processing. Experiments have shown that it has high detection accuracy and meets the needs of real-time engineering.

Acknowledgment. Supported by the National Key R & D Program of China (2018AAA0101703) and the key research and development project of Shandong province (2019GNC106093).

References

1. Lei, Y., Shao-yun, W., Li-ran, L., et al.: A pedestrian detection method in intelligent video monitoring system. Comput. Mod. (11), 69 (2019)
2. Raheja, J.L., Deora, S., Chaudhary, A.: Cross border intruder detection in hilly terrain in dark environment. Optik **127**(2), 535–538 (2016)
3. Li, G., Yang, Y., Qu, X.: Deep learning approaches on pedestrian detection in hazy weather. IEEE Trans. Ind. Electron. **67**, 8889–8899 (2019)
4. Dong, G., Song, C.-L.: Video-based pedestrian crossing detection system in mines. Ind. Mine Autom. (02), 29–34 (2017)
5. Cao, Z., Simon, T., Wei, S.E., et al.: Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291–7299 (2017)
6. Li, Q., Li, R., Ji, K., et al.: Kalman filter and its application. In: 2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS), pp. 74–77. IEEE (2015)
7. Papandreou, G., et al.: Towards accurate multi-person pose estimation in the wild, pp. 3711–3719 (2017). https://doi.org/10.1109/CVPR.2017.395.
8. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2