People Trajectory Forecasting and Collision Avoidance in First-Person Viewpoint

Guan-Yu Lai, Kuan-Hung Chen, and Bau-Jy Liang
Department of Electrical Engineering, Feng Chia University
Department of Electronic Engineering, Feng Chia University
apple03090@gmail.com, and kuanhung@fcu.edu.tw

***Abstract*—we propose a new collision avoidance system for first-person viewpoint, to show trajectory of people and to predict the future location of them. Then, we can determine the predicted location to avoid collision. We use deep learning to detect pedestrians and plot out coordinates of the trajectory. We predict future location of the target according to the law of inertia. In the first-person screen, this system can show whether possible collision occurs. Experimental results show that our method is feasible and outperforms state-of-the-art.**

Keywords—people recognition; predict track; Collision detection

I. INTRODUCTION

With the advancement of technology, the trend to make our life more convenient by automate technology is unstoppable. In the future, lot of automation technology will enter our living environments to do all kind of tasks, for example, the growing trend of self-driving car. Considering the society aging issue, the self-driving scooters can provide people necessary mobility. However, issues including how to avoid collisions with people in crowded environment need to be addressed.

Therefore, in the paper, we propose a new collision avoidance system for first-person viewpoint which can be divided into three steps, i.e., people recognition, people trajectory prediction and collision avoidance. For people recognition, we use a deep learning method to identify people in the scene. With other advanced object recognition system technologies such as R-CNN [3], Fast R-CNN [4], and faster R-CNN [5], in terms of real-time performance, YOLO [1] deep convolution neural networks, identified on the GPU can get in excess of 40 fps and also can get 95% accuracy. So, we use YOLO model to identify targets. Second, for the prediction of people trajectory, paper [6] trains a deep learning model based on three input information including local & scale stream, ego-motion stream and pose stream, to learn the prediction of people trajectory. Nevertheless, the calculation is relatively high. To solve this problem, we found that in general, people walk at an equal pace, so we can use law of inertia to predict the future location of people. The calculation complexity of the inertia is rather low. At last, for collision detection, we screen the dangerous areas in user frame. When the predicted location of people enters into the danger area, the system gives alerts.

In order to achieve the purpose, we modify the deep learning code of YOLO [2]. Our flow chart is shown in Fig. 1, by using deep learning to recognize people in the scene with high accuracy, and then predicting the future position based on the past trajectory of the people. With that, we can get high accuracy and avoid possible collision.

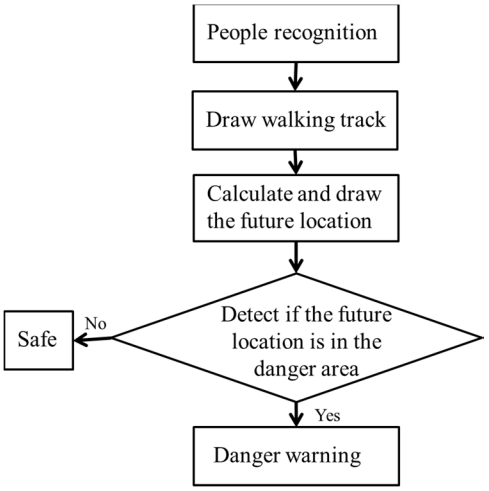


Fig. 1 The flow chart of the whole system

II. PEOPLE RECOGNITION

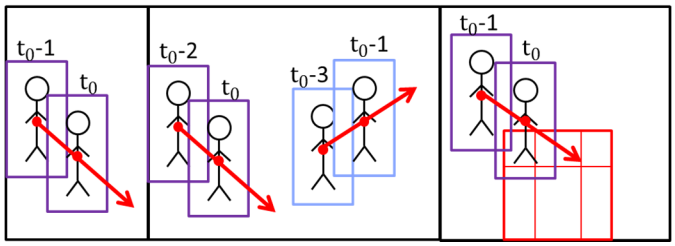
To recognize people accurately, we train a deep neural network model with 2012 Pascal VOC dataset. We only choose pictures with person and get the person label information. The training machine is GPU 1060; the adopted YOLO convolution neural network is yolo-voc.cfg [2]. We spent twelve hours for training our model, and get 95% accuracy on people recognition.

III. PROPOSED PREDICTION OF PEOPLE TRAJECTORY AND COLLISION AVOIDANCE

With YOLOv2 [2] recognized box, as illustrated in Fig. 2 (a), we can get the box quarter point and set it as a representative as the box. Then, by connecting the point, we can get the people movement track. In the general situation, people move at a constant speed, so using the people past position, we can predict the people future position by law of inertia. However, when we detect multiple people, each box has each time track as show in Fig. 2 (b). In the default displaying way of YOLO, only one box is plotted for each time instance. That is the boxes for the two people are displayed in turn. Hence, we can use the total number of targets, say n, the number of each box, namely m, the coordinate of box at time t, i.e., P_t to determine the predicted coordinate, as shown in equation (1).

$$P_{t-m+n} = 2P_{t-m} - P_{t-m-n},\ m = 0, 1, \ldots, n-1 \quad (1)$$

To avoid collision, we set a potentially dangerous area in the first person viewpoint and then screen whether there is any object inside this area. When predicted location of the people goes into the dangerous area, the area will show a warning box.



(a) (b) (c)

Fig. 2 (a) The prediction of people trajectory (b) Each box has each time track (c) Collision detection.

Besides, we also divide the area into up, left, right, top right, top left and middle sub-parts, as shown in Fig. 2 (c), to further classify the dangerous level of the obtrusive object.

IV. Experimental Results

We test all kinds of interaction behaviors, as defined in [6], i.e., turn around, across, toward and away. The corresponding demo result is shown in Fig. 3. The central bottom box is the pre-defined danger area. Solid lines indicate past ground truth tracks, and dotted lines represent predicted track. Fig. 3 (a) shows the case of a person who suddenly turns around. Although this situation is tough for prediction, we found that the predict track by using our algorithm can fit to the correct track. Along with Fig. 3 (a), Fig. 3 (b), (c), (d), (e) together show that our method is feasible for all kinds of interaction behaviors. Table1 shows the Euclidean distance average value of ground truth and predicted track of both [6] and our work. Because paper [6] did not release it method, so we can only report the results on the paper in Table1. Table 1 illustrates that our work outperforms [6] by obvious improvement in terms of Euclidean distance. Only our test data is not as many as that of [6] due to the limited open database as well as limited time. We will collect more test video and complete the test in recent future.

Furthermore, we found that the distance between us and each person is also significant. Hence, we add the predicted distance, as show in Fig. 4 (b), by using box area. Nevertheless, wrong prediction may occur if box is at the border of the view field. Accordingly, we have a calibration for the distance. Fig. 4 (a) shows the experiment place, which we can use to provide one more determination of distance to determine if there is danger.

V. Conclusion

In this paper, we designed a pedestrian collision avoidance system based on a deep learning neural network, i.e., YOLOv2 [2]. By using YOLOv2 [2] and law of inertial, we can predict track and enable collision avoidance no matter when surrounding people across our path or suddenly turn around. In the first person viewpoint, we also can know how far the people in front of us is. The preliminary experimental results show that our work outperforms state-of-the-art in terms of obviously more accurate prediction results.

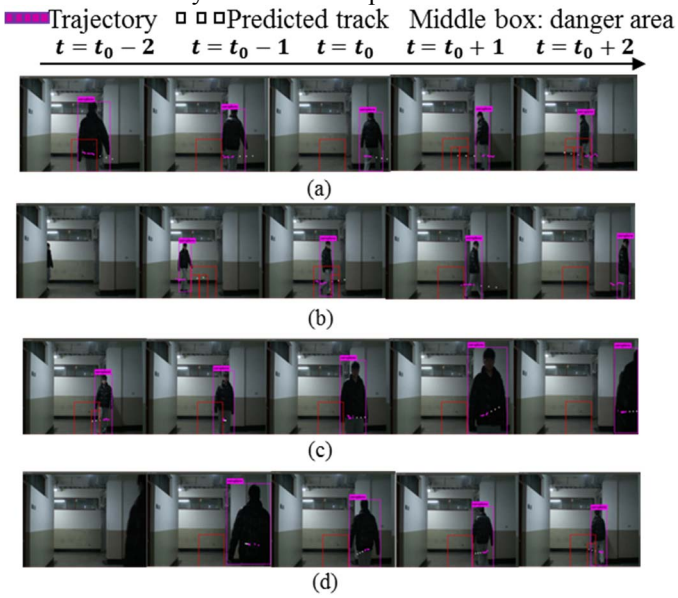
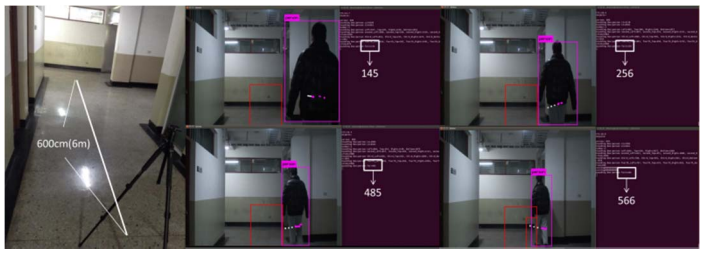


Fig.3 Demo result (a) Suddenly turn around (b) Across (c) Toward (d) Away



(a) (b)

Fig.4 (a) Experiment place (b) Add distance (cm) information.

TABLE1

Result with my demo video with Euclidean distance average value of predict track and ground truth, and shows the paper [6] result.

Behavior	Turn around	Across	Toward	Away
Paper [6]	----------	112.88	131.94	125.48
Our method	91.03	82.60	65.10	66.95

References

[1] J. Redmon, et al., "You only look once: Unified, real-time object recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016

[2] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 25 Dec 2016

[3] R. Girshick, et al., "Rich feature hierarchies for accurate object recognition and semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition. 2014

[4] R. Girshick, "Fast r-cnn," Proceedings of the IEEE international conference on computer vision. 2015

[5] S. Ren et al., "Faster R-CNN: Towards real-time object recognition with region proposal networks,"Advances in neural information processing systems. 2015

[6] Takuma Yagi, Karttikeya Mangalam, "Future Person Localization in First-Person Videos," Submitted on:Computer Vision and Pattern Recognition (cs.CV) 30 Nov 2017