Multi-feature and Modular Pedestrian Intention Prediction using a Monocular Camera

Mostafa Waleed and Amr EL Mougy
German University in Cairo, Cairo, Egypt

Keywords: Intention Prediction, Autonomous Vehicles, Advanced Driver Assistance System.

Abstract: Accurate prediction of the intention of pedestrians to cross the path of vehicles is highly important to ensure their safety. The accuracy of these intention prediction systems is dependent on the recognition of several pedestrian-related features such as body pose, head pose, pedestrian speed, and passing direction, as well as accurate analysis of the developing traffic situation. Previous research efforts often focus only on a subset of these features, therefore producing inaccurate or incomplete results. Accordingly, this paper presents a comprehensive model for pedestrian intention prediction that incorporates the recognition of all the above features. We also adopt the Constant Velocity Model to estimate the future positions of pedestrians as early as possible. Our model includes a reasoning engine that produces a decision based on the output of the recognition systems of all the aforementioned features. We also consider occlusion scenarios that happen when multiple pedestrians are crossing simultaneously from the same or different directions. Our model is tested on well-known datasets as well as a real autonomous vehicle, and the results show high accuracy in predicting the intention of pedestrians in different scenarios, including ones with occlusion among pedestrians.

1 INTRODUCTION

According to the World Health Organization (who, 7 30), nearly 1.35 million people die each year and 20-50 million people are injured as a result of road traffic accidents. More than half of all road traffic deaths are among vulnerable road users: pedestrians, cyclists, and motorcyclists.

Autonomous vehicles will have to consider the random actions of pedestrians to ensure their safety. Accordingly, the research community has been actively investigating methods of predicting as early and accurately as possible the intention of the pedestrian to cross the path of the vehicle. To achieve this objective, researchers have developed recognition systems for several pedestrian features to predict whether or not a pedestrian intends to cross. For example, body pose is used to detect the readiness of pedestrians to cross, head pose is used to detect if a pedestrian is aware of the incoming traffic, pedestrian direction and speed are used to detect if the path of a pedestrian will cross that of the vehicle.

Even though the existing recognition systems may be able to detect the above features accurately, intention prediction systems that are built on top of these recognition systems often fail to produce accurate results. This is mainly because considering only a subset of these features may not fully capture the intention of the pedestrian to cross.

Therefore, this paper presents a comprehensive pedestrian intention prediction model that simultaneously considers all the aforementioned features. We utilize computer vision and machine learning approaches for the recognition of each feature independently, and propose a weighted reasoning engine that combines the output of all the recognition modules to produce a prediction of the pedestrian's intention. In this reasoning engine, we specify proper weights that should be give to the output of each recognition module, according to its contribution to the intention prediction process. In addition, we address a critical challenge in the intention prediction which is the occlusion of pedestrians when there is more than one of them crossing. We propose a technique of frame analysis to detect occlusion and re-identify the pedestrians as the occlusion fades. In order to evaluate our approach, we test it in three different environments. First we evaluate our approach on our proposed dataset then we used KITTI dataset (Geiger et al., 2013) for more test cases. Finally we used our modified autonomous golf cart to evaluate our approach. Our proposed model shows promising results

Waleed, M. and El Mougy, A.
Multi-feature and Modular Pedestrian Intention Prediction using a Monocular Camera.
DOI: 10.5220/0010337711601167
In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021) - Volume 2*, pages 1160-1167
ISBN: 978-989-758-484-8

in all cases. The remaining sections of this paper are organized as follows. Section 2 reviews pertaining research efforts. Section 3 offers the full details of our proposed model. Section 4 shows the results of the performance evaluation, while Section 5 provides concluding remarks.

2 RELATED WORK

In this section we discuss the main recent contributions in pedestrian intention prediction. The authors of (Joon-Young Kwak, 2017) propose a dynamic fuzzy automata (DFA) method for pedestrian intention and use low-level features with a boosted-type random forest classifier for pedestrian detection and tracking. To consider the pedestrian characteristics they use the pedestrian's distance from curb , pedestrian's speed and the direction of his/her head. Four pedestrian intention states are defined, two of them represent that the pedestrian is not passing, and the other two represent that the pedestrian is passing. This approach has 9 FPS processing time which is not sufficient for real time applications. In another paper, Sebastian Kohler et al.(S. Köhler and Dietmayer, 2012) show that the sub-region of the image that covers the pedestrian within its bounding box is available for a time series, e.g. by the fusion of LIDAR and video-data and a HOG-based detection. The methodology of their approach is to generate the motion descriptors within this box and to classify the motion. This approach was tested on lab conditions so it wasn't proven yet it's efficiency in real time applications. Gurkan Solmaz et al.(Solmaz et al., 2019) propose the use of Internet Of Things (IOT) technology where the pedestrian next location is predicted based on his/ her historical data and current position. Both the pedestrian GPS position and velocity are obtained using a mobile device to predict the pedestrian's next position using a trajectory model. This approach assumes that all the pedestrians are using a 4G mobile device and that pedestrians are always walking in the same direction, which is not the case. Christoph Scholler et al.(C. Schöller and Knoll, 2020) used a simple constant velocity model(CVM) to predict the pedestrian intention that does not require any information besides the pedestrian's last relative motion. They denote the position $({x_t}^i, {y_t}^i)$ of pedestrian i at time-step t as ${P_t}^i$. The goal of pedestrian motion prediction is to predict the future trajectory Ti = $({P_t}^{i+1}, \ldots.., {P_t}^{t+n})$ for pedestrian i, taking into account his or her own motion history $H_i = ({p_0}^i, \ldots, {p_t}^i)$.The constant velocity model approach mispredict the pedestrian intention if he/she suddenly change his/her walking direction. In (Rehder et al., 2018), the authors propose a different approach that relies on predicting pedestrian intention using goal directed planning. They use a mixture density function for possible destinations. They use these set of destinations as the goal states of a planning stage that predict the motion of the pedestrian based on the common motion patterns that are already known. Those patterns are learned by a fully convolutional network operating on the maps on the environment. R. Quintero et al .(Quintero et al., 2014) considered the three-dimensional pedestrian body language in order to perform path prediction in a probabilistic framework. For this purpose, the different body parts and joints are detected using stereo Vision. The body pose algorithm they use predict the input as a point cloud on one pedestrian that has been previously extracted from the general point cloud provided by the stereo images pair. Let P = $\{p_1, ..., p_N\}$ represent the pedestrian point cloud with N points.The recursive nature of the algorithm limits the accuracy of a body part on the accuracy of the previous part. If a part is incorrectly detected all following parts will be affected. In our proposed approach we overcame all the mentioned limitations by using multi features in order to ensure the prediction results and we considered the processing time to be able to fit our approach in real time applications.

3 INTENTION PREDICTION MODEL

Our system architecture consists of several stages. First, a frame captured by the monocular camera acts as input to our system. Then, the human detection model and the head pose model use the captured frame as an input. The output for the human detection model is a bounding box for each pedestrian in the frame which is used as the current location for the pedestrian while the head pose model output the head orientation for all the pedestrians. The frame with bounding boxes after pre-processing act as input for the body pose estimation model and constant velocity model while the bounding box points are used to detect the pedestrian direction and the side from which the pedestrian will pass. Also the pedestrian position is used to detect their moving speed. The output of the system is the person's future position predicted using the constant velocity model and the person's intention as "passing" or "Not passing".

3.1 Pedestrian Detection

For pedestrian detection, we use real-time object detection model YOLO proposed in (Redmon and Farhadi, 2018). The input for YOLO v3 is a (720 x 1280 x 3) frame and the output is a (720 x 1280 x 3) frame where a bounding box marks each person in the frame in 2D space which is in form of (x,y,width,height) for each pedestrian. The position of pedestrian n is defined as $\{CenterX_n, CenterY_n\}$ where $CenterX_n = x_n + (w_n/2)$ and $CenterY_n = y_n + (h_n/2)$

3.2 Pedestrian Tracking

After detecting the pedestrians, we need to track them through the frames as we need the pedestrian history to be able to predict his/her next move. So, to track the pedestrian n between two frames at time t-1 and t we calculate the Euclidean distance between the center of the pedestrian $Center_n^{t-1}$ at t-1 and all the appearing centers of pedestrians at time t. So, we have the centers of the pedestrians at time t as $\{Center_1^t, ..., Center_n^t\}$. Thus, to match pedestrian n between two frames we use the following equation:

$$\min\left(\sqrt{(|CenterX_n^t - CenterX_S^{t-1}|)^2 + (|CenterY_n^t - CenterY_S^{t-1}|)^2}\right) 1 \leqslant S \leqslant N \quad (1)$$

A problem was encountered that when occlusion occurs between pedestrians our tracking technique mismatches those pedestrians. Accordingly , when a pedestrian is not detected for one or more frames we match the lost pedestrian based on the body features (e.g. body height , body width) and passing direction, not only the euclidean distance.

3.3 Features Extraction

3.3.1 Pedestrian Passing Direction (PPD)

The PPD provides a meaningful cue for prediction of pedestrian intention. If the pedestrian is on the right side he/she needs to move in the negative direction of X axis as shown in Fig.1 to cross through the car path. PPD helps to know if the pedestrian is moving in the direction of the car or in the other direction. Thus, the other extracted features can depend on the PPD to know if the pedestrian is passing or not.

In order detect the PPD and avoid the fluctuation in the pedestrian's bounding box we use a six state Mealy finite state machine (FSM) shown in Fig.2. Each pedestrian in the frame has his own PPD-FSM where, if the pedestrian is passing from the left side to the right side, the FSM will be in one of the three states(left-side0, left-side1, left-side2). The

Figure 1: PPD axis definition.

left-side0 state is the state with highest confidence that the pedestrian's direction is from left to right. Alternatively, if the pedestrian is passing from the right side to the left side the FSM will be in one of the other three states (right-side3, right-side4, right-side5), where right-side5 state is the state with highest confidence that the pedestrian's direction is from right to left.

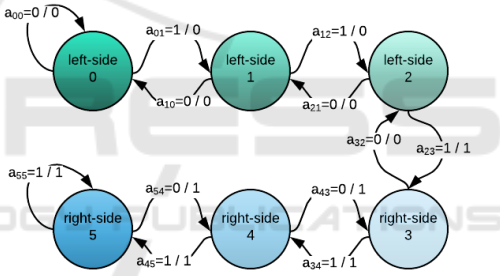

Figure 2: The Finite State Machine of the PPD module.

In the first Frame for pedestrian n the PPD-FSM start state is determined based on the pedestrian position relative to the car. As the camera position is the center of the frame, we can conclude that the car is at the camera position. So, if the position of pedestrian n in the X-axis is less than that of the the camera position, the FSM start state will be left-side0. If pedestrian n position in the X-axis in more than that of the camera position, the FSM start state will be left-side5. The PPD-FSM for pedestrian n is updated every 5 consecutive frames. So, to know the direction of pedestrian n we use his/her position at time t $\{CenterX_n^t, CenterY_n^t\}$ and time t-5 $\{CenterX_n^{t-5}, CenterY_n^{t-5}\}$.

3.3.2 Pedestrian Moving Speed and Direction (PMSD)

The PMSD of a pedestrian offers an important clue for estimating his/her action. The objective here is to

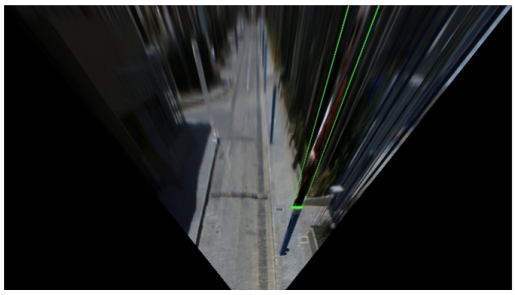

Figure 3: Orthogonal image.

detect whether the pedestrian is moving in the X direction or in the Z direction. The problem with images that all the lines are collected at one vanishing point, so even if the pedestrian is moving parallel to the car path(in the Z direction) in the real world he/she will appear moving in the car path (in the X direction) in the image not only the Z direction due to the convergence from 3D space to 2D space. Thus, to be able to calculate the pedestrian speed and determine the direction accurately, we need to remove the perspective incurred by the fact that the camera is mounted on the car. We convert the perspective image to an orthogonal image that allows the perspective effect to be removed from an image as shown in Fig.3.

After converting the original image to an orthogonal image we can extract the new position of the pedestrian after removing perspective effect. Accordingly, we crop the bounding box of each pedestrian from the the original image shown in Fig.3 . Then, we convert the image for each pedestrian to an orthogonal image. The new position in the X axis now can be calculated using the color of the bounding box. So, the new position of pedestrian n at time t is $\{NCenterX_n^t, CenterY_n^t\}$. To be able to detect the PDMS we need to calculate the deltas between the pedestrian new position at time t and time t-1 which is defined as :

$$delta_x = NCenterX_n^t - NCenterX_n^{t-1} \quad (2)$$

$$delta_y = CenterY_n^t - CenterY_n^{t-1} \quad (3)$$

$delta_x$ is the difference between the X positions in two consecutive frames so it can be considered as the velocity in X direction (vel_x), and $delta_y$ can also be considered the velocity in Y direction (vel_y). To determine the pedestrian movement direction we faced a problem that the pedestrian bounding box position sometimes fluctuates. To solve this problem, we use seven state Mealy finite state machine to detect pedestrian state (PS-FSM), as shown in Fig.4. This way, we can maintain the correct direction even if the bounding box fluctuates in one of the frames. The PS-FSM state zero (Start state 0) is a one time visit start state. In the first frame, the pedestrian position is determined if he/she is on the left side of the car or right side on the car using the PPD-FSM. If the pedestrian is on the left side or the right and walking toward the car direction, PS-FSM will be in one of the three states(passing1, passing2, passing3). But if the pedestrian is walking in X direction away from the car or the pedestrian is walking in the Z direction, the PS-FSM will be in one of the three states (Notpassing4, Notpassing5, Notpassing6). The state transition (event) is represented by arcs between the nodes, as shown in Fig.4, and the state transition from a state i to a state j is described as b_{i_j}.

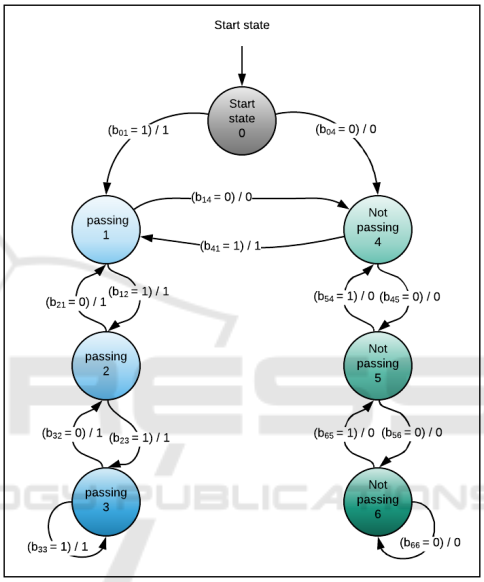

Figure 4: Pedestrian state finite state machine(PS-FSM).

3.3.3 Head Orientation (HO)

When pedestrians are moving, their HO tends to coincide with the direction of their movement. Therefore, we can predict the moving direction of the pedestrian if we can estimate his/her HO. HO estimation consists of face detection and head orientation estimation. First, for face detection we use a pre-trained model[1] on a data set generated by using the cleaned widerface labels. The head detection model take as an input a frame that contains the pedestrian and the output is a bounding box points for each pedestrian face in the frame as shown in Fig.5.

After detecting the face position, the head orientation can be predicted using Hopenet pre-trained model proposed in (Ruiz et al., 2018) that takes as in-

[1]https://github.com/Linzaer/Ultra-Light-Fast-Generic-Face-Detector-1MB

put a cropped image of the pedestrian face using the bounding box of the face and outputs the predicted pedestrian yaw, roll and pitch as shown in Fig.5.

After extracting all the HO of all pedestrians in a frame, we start matching those head poses with the detected pedestrians. The matching is done by using the face bounding box and the person bounding box. We are mainly interested in the pedestrian's head yaw value as it indicates the rotation angle of the head around the Y axis.

Figure 5: Head orientation estimation.

3.3.4 Body Pose (BP)

Pedestrian BP is a very important clue that can predict if this person is going to pass or not even if the pedestrian position is still constant. Before the pedestrian starts using his/her legs to walk, he/she first bends his/her upper body to give himself/herself the first push to start the motion. Thus, we use the BP to be able to predict the pedestrian's next move. To recognize the BP we use a pre-trained model proposed in (Cao et al., 2017). The system takes as input a color image of size $w \times h$ and produces the 2D locations of anatomical keypoints for each person in the image as shown in Fig.6.

To be able to predict the pedestrian's next move we were interested in some keypoints such as right shoulder, left shoulder, nose, right knee, left knee, left ear, and right ear. The left and right shoulder were the main indicators that the upper body is bending while the left and right knees were the indicators that the pedestrian started moving. To be able to detect changes in keypoints, we calculated the deltas between each keypoint at time t-1 and its corresponding keypoint at time t. For example, if the right knee position at time t-1 is $\{Rknee_x^{t-1}, Rknee_y^{t-1}\}$ and the right knee position at time t is $\{Rknee_x^t, Rknee_y^t\}$, the deltas of the right knee $delta_{Rknee}$ is defined as:

Figure 6: Image with pedestrian BP.

$$delta_{Rknee} = \{|Rknee_x^t - Rknee_x^{t-1}|, |Rknee_y^t - Rknee_y^{t-1}|\} \quad (4)$$

Each body part (keypoint) detected has its own confidence based on how effective this part can affect the pedestrian movement. Left and right knees and left and right shoulders have the highest confidence as they are the main contributors to predict if this pedestrian is going to pass or not. So, for each of the keypoints, a $delta_{bodypart}$ is calculated between two frames and if this delta is greater than a certain *threshold*, the passing confidence of this part is added to the total confidence, if the delta is in the direction of the car that is defined using the PPD-FSM mentioned in 3.3.1. Also, backup deltas are computed for each keypoint but only every 3 frames. For example, if the right knee position at time t-3 is $\{Rknee_x^{t-3}, Rknee_y^{t-3}\}$ and the right knee position at time t is $\{Rknee_x^t, Rknee_y^t\}$, the back up delta of the right knee $backupdelta_{Rknee}$ is calculated as:

$$backupdelta_{Rknee} = \{|Rknee_x^t - Rknee_x^{t-3}|, |Rknee_y^t - Rknee_y^{t-3}|\} \quad (5)$$

These back up deltas help to detect small range body movements to be able to predict the pedestrian intention as early as possible. Each keypoint delta or backup delta has its own *threshold* that is picked based on the pedestrian position in the Y axis since the pedestrian change in X position differ based on the position in Y position. Thus, we divided the image into 10 regions to define the thresholds for different body parts. Each of these regions has its own average and minimum deltas for each body part.

3.4 Constant Velocity Model (CVM)

After extracting all the features mentioned in the previous section (PPD, PMSD, HO, BP) we can predict pedestrian's intention. However, this is not the only thing we need to know about the pedestrian. Self driving cars also need to know the next position of each pedestrian in order to handle safe maneuvers and stops. To predict the pedestrian's future position we use CVM approach. CVM uses the motion history of the pedestrian to be able to predict his/her next position. We donate the position $\{NCenterX_n^t, CenterY_n^t\}$ of pedestrian n at time t as p_n^t, which is extracted from the orthogonal image as mentioned in section 3.3.2. The goal of pedestrian motion prediction is to predict the future trajectory $F_n = (p_n^{t+1}, ..., p_n^{t+m})$ for pedestrian n, taking into account his/her motion history $H_n = (p_n^0, ..., p_n^t)$. We use average displacements between the last 3 frames to be able to predict the future displacement, average displacement, defined as:

$$D_{avg} = \frac{(p_n^{t-2} - p_n^{t-3}) + (p_n^{t-1} - p_n^{t-2}) + (p_n^t - p_n^{t-1})}{3} \quad (6)$$

So in order to predict the future position, we use D_{avg} as a constant velocity for the pedestrian n in the next m frames. So p_n^{t+1} is calculated by adding the constant velocity of the pedestrian n D_{avg} with p_n^t so the future position F_n, defined as :

$$F_n = (p_n^t + D_{avg}, ..., p_n^{t+m-1} + D_{avg}) \quad (7)$$

3.5 Final Prediction

The final prediction is obtained based on the output of the prediction modules (BP, PMSD, CVM, and HO). Each of these modules can predict that pedestrian n is passing with a certain percentage. Then, we apply a weighted sum to obtain the final prediction. This weighted sum is defined as :

$$Final\ prediction = \sum_{i=1}^{4} module_i * confidence_i \quad (8)$$

Where $module_i$ is the output percentage of one of the prediction modules and the $confidence_i$ is how much we trust the output of this module. In this paper, $confidence_i$ is obtained based on the error rate of each module which is calculated in Section 4.

4 PERFORMANCE EVALUATION

In order to compare our proposed approach to the other approaches, we use six videos from KITTI datasets (Geiger et al., 2013) to evaluate our own model. The six videos contains in total 227 frames and 13 pedestrians. We also propose a novel dataset to evaluate pedestrian intention prediction. The proposed dataset is captured in the German University in Cairo(GUC) and consists of eight 2048×1536 sized videos captured by a camera mounted on the front-top of a moving car.Where the 8 videos of the proposed dataset contains in total 1123 frames and 20 pedestrians. We use the videos in the two datasets as input for our system. For each video, we evaluate our approach based on the output for each frame. Thus, we evaluate Body Pose and Head Pose detection and orientation estimation, the Pedestrian Moving Speed and Direction, the pedestrian Passing Direction, and the prediction of the constant velocity model accuracy. We also test the proposed approach on our self-driving car to make sure that our approach is compatible for real time applications.

4.1 Feature Descriptors for Pedestrian Intention Prediction

In order to evaluate the proposed approach, we labeled the GUC dataset for each pedestrian appearing in each video. Thus, we classify each pedestrian state in each frame as "Passing" or "Not passing" to be able to evaluate our proposed approach on the proposed dataset. Then, we calculate the error rate of the following individual features:(1)PMSD, (2)BP, (3)HO, (4)CVM, and (5)PPD, using the following equation :

$$Error\ rate = \frac{Faulty\ Predictions}{Total\ Number\ of\ Predictions} \quad (9)$$

As shown in Fig.7, HO exhibits the lowest performance with respect to error rate as 87 %. The major reason for the lowest performance of HO is the false decision of the face detection and orientation estimation when the pedestrian is located a long distance away from the camera. In contrast, PDMS shows lower error rate of 18% . According to the error rate graph, we found that some features cannot be used alone to predict the pedestrian intention (such as HO and the BP) to avoid false predictions. Also PPD, which has the lowest error rate of 2%, cannot be used to predict the pedestrian intention as it only tells us which curb the pedestrian is passing from. Thus, in our proposed approach we combine all these individual features(PDMS+BP+HO+PPD+CVM) to predict the pedestrian intention and future position. The pedestrian intention is detected using BP+HO+PDMS+PPD and future position is predicted using CVM that has error rate of 25.8%.

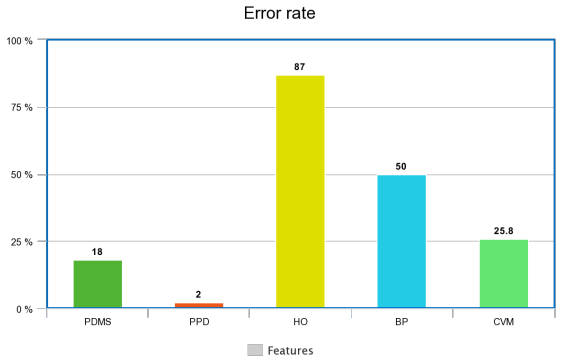

Figure 7: Performance comparison between five individual features.

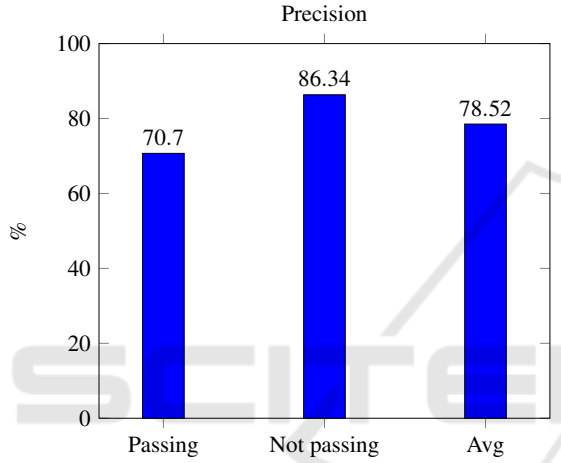

Figure 8: Precision.

4.2 Performance of Pedestrian Intention Prediction

For performance evaluation of pedestrian intention prediction, we evaluate the precision for our proposed approach. We calculate the precision of the two output classes for our proposed system which are: Passing and Not passing as shown in Fig.8. Precision (P) is defined as the number of true positives (T_P) over the number of true positives plus the number of false positives (F_P) as shown in Equation.10.

$$P = \frac{T_P}{T_P + F_P} \tag{10}$$

As shown in Fig.8 our proposed approach reaches a very good precision percentage of 86.34% to predict if this pedestrian is Not passing and reaches about 70.7% to predict that this pedestrian is passing. The reason why the precision of Passing is less than that of Not passing is that our approach needs about two or three frames to be able to predict that this pedestrian is changing his/her state from Not passing to Passing. So, in average the precision was found to be 78.52%.

4.3 Processing Time of the Pedestrian Intention Prediction

In addition to the prediction performance, we were concerned with the computational speed of the proposed approach as it should work in real-time. It was found that the processing time of the proposed approach is 11 FPS on average which is sufficient for real-time applications as KITTI dataset (Geiger et al., 2013) which is used for testing is captured at 10 FPS.

4.4 Pedestrian Intention Prediction Results

In this section we will show the results of our proposed approach on two videos of the proposed dataset (GUC dataset), KITTI dataset and a live test on our self-driving car. In these results we will present different scenarios for pedestrians standing, walking parallel to the car direction, and passing in front of the car.

4.4.1 GUC Dataset

Fig.9 shows the result of the pedestrian intention prediction of the proposed algorithm for five of the test videos of the GUC dataset. As shown in Fig.9 (a) the pedestrian actions is classified into two categories: ”Passing” and ”Not passing”. The examples shown in Fig.9 shows good performance of the proposed algorithm for pedestrian intention. In Fig.9 (a)(2) our approach mispredicted the intention of the pedestrian as the pedestrian was not appearing due to the trash bin and then a recovery happens in Fig.9 (a)(3) and the prediction was adjusted.

4.4.2 KITTI Dataset

Fig.9(b) shows the result of the pedestrian intention prediction of the proposed algorithm on one of the test videos of KITTI dataset. As shown in Fig.9 the pedestrian’s actions are classified into two categories: ”Passing” and ”Not passing”. The examples shown in Fig.9 (b) shows good performance of the proposed algorithm for pedestrian intention. In Fig.9 (b) (1) a pedestrian is passing and is matched correctly. However, In Fig.9 (b) (2) mismatching happens and the passing pedestrian appeared to be Not passing. In Fig.9 (b) (3) our approach recovered and rematched the pedestrian and predicted that he is passing again correctly.

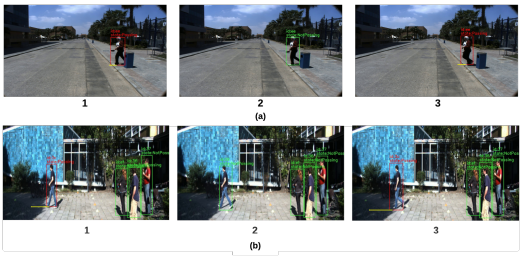

Figure 9: Results for pedestrians' intention prediction (GUC & KITTI dataset).

4.4.3 Live Testing

In this section we are going to show the results of the proposed approach and how it was working in real time scenarios on our modified self-driving vehicle. We tested our system on our self-driving car "Herbie" within the campus of the German International University(GIU). The test cases were a pedestrian walking in front of the car and pedestrian running in front of the car and a pedestrian is passing and intersects with the car path. Our system detects the pedestrian and outputs that there is a pedestrian passing in front of the car and the system on the car made the appropriate action and stopped.

5 CONCLUSIONS

In this paper, we propose pedestrian intention prediction system based on machine learning and computer vision. To predict if the pedestrian is passing or not HO, BP, PDMS, PPD features were extracted. We also use CVM to be able to predict the pedestrian's future position so we can avoid the collision between the car and the pedestrian. Our system is evaluated using datasets and live testing, where it is proven that our approach is effective in detecting pedestrian intention and future position. Also, the performance of our proposed approach was found to be sufficient for real time applications. In future research, we will focus on compensating the ego-motion on the camera to be able to remove the effect of car movement, as the pedestrian's motion is influenced by the distance between the pedestrian and the car. Furthermore, we will use more accurate and advanced matching and tracking techniques to avoid mis-predictions due to false matches.

REFERENCES

(accessed 2020-07-30). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

C. Schöller, V. Aravantinos, F. L. and Knoll, A. (2020). What the Constant Velocity Model Can Teach Us About Pedestrian Motion Prediction. 5.

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Real-time multi-person 2d pose estimation using part affinity fields. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Geiger, A., Lenz, P., Stiller, C., and Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*.

Joon-Young Kwak, Byoung Chul Ko, J.-Y. N. (2017). Pedestrian intention prediction based on dynamic fuzzy automata for vehicle driving at nighttime.

Quintero, R., Almeida, J., Llorca, D. F., and Sotelo, M. A. (2014). Pedestrian path prediction using body language traits. pages 317–323.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*.

Rehder, E., Wirth, F., Lauer, M., and Stiller, C. (2018). Pedestrian prediction by planning using deep neural networks.

Ruiz, N., Chong, E., and Rehg, J. M. (2018). Fine-grained head pose estimation without keypoints. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

S. Köhler, M. Goldhammer, S. B. K. D. U. B. and Dietmayer, K. (2012). An analysis of the current state of English majors' BA thesis writing [J]. *Early detection of the Pedestrian's intention to cross the street*, 3.

Solmaz, G., Berz, E. L., Dolatabadi, M. F., Aytaç, S., Fürst, J., Cheng, B., and den Ouden, J. N. (2019). Learn from iot: Pedestrian detection and intention prediction for autonomous driving. In *SMAS '19*.