Understanding Pedestrian Behavior in Complex Traffic Scenes

Amir Rasouli, Iuliiia Kotseruba, and John K. Tsotsos

Abstract—Designing autonomous vehicles for urban environments remains an unresolved problem. One major dilemma faced by autonomous cars is understanding the intention of other road users and communicating with them.

To investigate one aspect of this, specifically pedestrian crossing behavior, we have collected a large dataset of pedestrian samples at crosswalks under various conditions (e.g. weather) and in different types of roads. Using the data, we analyzed pedestrian behavior from two different perspectives: the way they communicate with drivers prior to crossing and the factors that influence their behavior.

Our study shows that changes in head orientation in the form of looking or glancing at the traffic is a strong indicator of crossing intention. We also found that context in the form of the properties of a crosswalk (e.g. its width), traffic dynamics (e.g. speed of the vehicles) as well as pedestrian demographics can alter pedestrian behavior after the initial intention of crossing has been displayed. Our findings suggest that the contextual elements can be interrelated, meaning that the presence of one factor may increase/decrease the influence of other factors. Overall, our work formulates the problem of pedestrian-driver interaction and sheds light on its complexity in typical traffic scenarios.

Index Terms—Autonomous driving, intelligent vehicles, driver-pedestrian interaction, human intention and behavior analysis, pedestrian behavior understanding, safety and collision avoidance.

I. INTRODUCTION

Autonomous driving technologies have come a long way since the first introduction of commercial automobiles in the 1920s. Ever since the early attempts on achieving autonomy [1], we have witnessed great success stories throughout the past decades such as autonomous driving on highways [2], platooning [3], unsupervised driving on rough terrains [4] and urban environments [5], and even autonomous racing cars [6]. Today, autonomous driving is one of the major topics in technology research and a large number of companies have been heavily investing in it. According to some economists, the size of the global autonomous vehicle industry and related software and hardware technologies is estimated to be more than 40 billion dollars by the year 2030 [7].

The current level of autonomy available on some commercial cars such as Tesla is level 2, and some manufacturers such as Audi are promising level 3 capability on their latest models such as A8 [8]. According to SAE standard, autonomy level 3 means that the car can handle all aspects of the dynamic driving task in specific environments with the exception that human driver may be requested to intervene [9]. Now the question is how far are we from achieving level 5 or fully autonomous behavior? The answer to this question is controversial. Some companies such as Tesla [10] and BMW [11] are more optimistic and claim that they will have their first fully autonomous vehicles entering the market by 2020 and 2022 respectively. Other companies such as Toyota are more skeptical and believe that we are nowhere close to achieving level 5 autonomy yet [12].

Aside from challenges associated with developing suitable infrastructure [13] and regulating autonomous cars [14], technologies currently used in autonomous vehicles have not achieved the level of robustness to handle various traffic scenarios such as different weather or lighting conditions (e.g. snow, rain), road types (e.g. driving on roads without clear marking or bridges) or environments (e.g. the GPS localization problem in cities with high-rise buildings) [15].

In addition, when it comes to driving in complex scenes, such as urban environments, autonomous vehicles face another dilemma, namely interacting with other road users [16]. In order to do so, the autonomous vehicle needs to understand their intentions, which can be achieved by communicating with them and predicting what they are going to do next (see Fig. 1). Interaction between traffic participants is vital for a number of reasons:

1) Ensures the flow of traffic. We as humans, in addition to official traffic laws, often rely on informal laws (or social norms) to interact with other road users. Such norms influence the way we perceive other road users and how we interpret their actions [17]. We also often engage in non-verbal communication to exchange our intentions in order to disambiguate certain situations. For instance, if a driver wants to turn onto a road without traffic signals, he/she might wait for another driver's
signal to indicate the right of way. Failure to understand others’ intentions can often result in accidents, which is evident from recent reports on Google’s autonomous test vehicles [18], [19].

2) **Increases safety.** Interaction also ensures the safety of road users, in particular, pedestrians as the most vulnerable traffic participants. For instance, at the point of crossing, pedestrians often establish eye contact with drivers or wait for an explicit signal from the drivers to ensure that they have been noticed and it is safe to cross [20].

3) **Prevents being subject to malicious behaviors.** Given that autonomous cars may potentially commute without any passengers on board, they may be subject to bullying [17]. For example, people might jump in front of the vehicles, forcing them to stop or change direction. Instances of bullying have been reported for the autonomous robots that are currently being used in malls. Some of these robots were defaced, kicked or pushed over by drunk pedestrians [21].

To address the above issues, one might argue that autonomous vehicles can be designed to be very cautious and conservative by always assuming the worst-case scenario and being prepared for it. However, it has been shown that conservative driving not only does not reduce accidents but can be very disruptive itself [22].

The automotive industry offers a number of solutions for dealing with the interaction problem. For instance, vehicle-to-vehicle (V2V) and vehicle to infrastructure (V2I) techniques that use cellular technology for communication are widely studied [23], [24]. Similar vehicle-to-everything (V2X) approaches also have been proposed for communicating with pedestrians [25]. In this method, the pedestrian’s phone broadcasts its position to warn the autonomous cars about the pedestrian’s presence. V2X technologies, however, raise a number of concerns. First, the communication is highly dependent on all components to function properly. A malfunction in any communication device in any of the sub-systems involved can lead to catastrophic safety issues. In addition, sharing information can raise some privacy concerns regarding the transmission of personal information. Last but not least, although these technologies can be effective in communicating dynamic information (e.g. velocity and position) of the road users, they cannot capture the higher level intention of the involved parties.

The aforementioned shortcomings point to the need for a new approach to resolving the interaction problem in traffic. In this paper, we address the issue of interaction between traffic participants, in particular, between drivers and pedestrians. For this purpose, we collected a large dataset consisting of videos of pedestrians crossing (or intending to cross) the street. We annotated this data with various behavioral and contextual tags in an attempt to identify factors that influence pedestrian decision-making at the point of crossing.

More specifically, in this work we address the following issues through the analysis of our dataset: identifying pedestrian crossing patterns from a moving car perspective, realizing how pedestrians communicate their intention, and finding various factors that influence pedestrian decision-making process.

II. RELATED WORKS

A. Communication in traffic

The role of non-verbal communication in resolving traffic ambiguities is emphasized by a number of scholars [26], [27], [28]. In traffic scenes, communication is particularly challenging because, first, there is no official set of signals and many of them are ambiguous, and second, the type of communication may change depending on the atmosphere of the traffic situation, e.g. city or country [29].

The lack of communication or miscommunication can greatly contribute to traffic conflicts. It has been shown that more than a quarter of traffic conflicts occur due to the absence of effective communication between road users. Out of the conflicts caused by miscommunication, 47% of the cases occurred with no communication, 11% were due to the ineffective communication and 42% happened during communication [29].

Traffic participants use different methods to communicate with each other. Such as [28] lists some of the means of communication used by road users. According to the author, pedestrians use eye contact (glancing/staring), handwave, smile or nod. Drivers, on the other hand, flash lights, wave hands or make eye contact.

In the context of autonomous driving, a number of scholars evaluate different means whereby an autonomous vehicle in the absence of the driver can communicate with pedestrians [30]. Methods such as using LED lights [31] or displays on the vehicle [32] have been investigated.

To interpret communication signals, one should take the context into account. For instance, pedestrians initiate eye contact to ask for the right of way whereas drivers establish eye contact to yield to the pedestrians. Handwave by a pedestrian may convey different meanings too, e.g. request for the right of way or signal to show gratitude. Therefore, to understand the intended meaning of a signal, it is important to know the context within which the signal is observed (see Fig. 2).

2a) Yielding, b) asking for the right of way, c) showing gratitude, and d) greeting a person on the other side of the street.
B. Context and pedestrian behavior understanding

In the psychology literature, pedestrian crossing behavior is linked to various factors among which the most influential ones are dynamic factors, social factors and physical context. Dynamic factors are related to the distance to approaching vehicles and their velocities. For instance, gap acceptance, or the time gap between vehicles, determines how safe pedestrians feel about crossing [33]. Gap acceptance is measured in terms of Time-To-Collision (TTC) or how far (in seconds) the approaching vehicles are from the point of impact [34]. Vehicle speed in isolation also can influence pedestrian visual perception. Studies show that the faster the speed of the vehicle, the less accurate is the pedestrian’s estimation of the vehicle’s speed [27].

Besides dynamics, social factors are found to have a great impact on pedestrian behavior. In particular, social norms determine the extent to which pedestrians obey the law [26], take risks [34] or the way they communicate with one another [17]. Another determining factor is pedestrian group size which influences pedestrians’ speed [35], the level of risk taking [36], or even the way drivers would react to it [37]. For instance, larger groups tend to reduce the speed of pedestrians. At the same time, pedestrians feel safer in large groups and are willing to accept a shorter time gap than when they are crossing individually [36].

Moreover, pedestrian behavior may vary in different physical contexts. For instance, a pedestrian who is crossing a signalized intersection will more likely cross without looking at the traffic because he/she expects the road users to comply with the signal. On the other hand, in the absence of a traffic signal, the pedestrian may be more conservative and will only cross the street if the approaching vehicles yield to him/her [26]. The geometrical features of the road also can impact crossing behavior. For example, pedestrians have a longer gap acceptance in wide streets compared to narrow ones [34].

Behavioral analysts have found factors such as demographics [38], pedestrian state [39] and traffic characteristics [40] to play a role in pedestrian decision-making. However, detailed discussion of these factors is beyond the scope of this paper.

The missing element in the pedestrian behavioral studies is establishing connection between different factors. The majority of the studies focus on an isolated issue such as the role of physical context or dynamics on pedestrian behavior and typically do not explain how these elements are connected to one another. Moreover, these studies generally focus on analyzing pedestrian behavior in a fixed or limited context, e.g. only at signalized intersections, fixed camera position, or specific groups of pedestrian such as children or the elderly.

C. Pedestrian intention estimation

In intelligent transportation research, the task of pedestrian intention estimation is generally treated as a tracking problem. This means that the algorithms developed for this purpose treat the pedestrian as a moving object and try to model its dynamic behavior to estimate its future location [41], [42]. Some works also consider the vehicle’s motion and distance to the pedestrian and show that better prediction results can be achieved [43]. A major drawback of these models is that they cannot predict non-motion, i.e. when the pedestrian stops the motion tracking prediction fails. This means that dynamics-based models are only effective when the motion is continuous.

Very few works have attempted to explicitly take advantage of contextual information for pedestrian path prediction. In some works, for instance, the authors assign a goal to each pedestrian based on which they predict the trajectory of the pedestrians [44]. Others look at head orientation as a sign of pedestrian awareness. They argue that when a pedestrian is looking towards the traffic, he/she is more likely to coordinate with the approaching vehicles prior to crossing [45].

A number of studies also use contextual elements such as group size [46] or street structure [47] for predicting pedestrian behavior. However, these works are done in simulation and none of them propose a practical solution to the visual perception problem for identifying contextual elements.

III. Method

A. Instrumentation

Three drivers over the age of 30 participated in our data collection procedure. The drivers utilized two SUV passenger cars (one small and one mid-size) during the recording sessions. The data was collected using three HD camera models, namely GoPro HERO+, Garmin GDR-35 and Highscreen Black Box Connect. In each recording session only one camera was used at a time, which was placed inside the car below the rear view mirror (see Fig. 3). This ensures that the camera remains inconspicuous to the eyes of pedestrians.

The data was collected in a naturalistic setting, meaning that recording took place as part of the participants’ daily activities and pedestrians were not notified about the recording. In addition, no particular instruction was given to the drivers for changing their driving habits.

B. The data

We recorded approximately 240 hours of driving footage over a period of 6 months in 5 geographical locations including Canada, USA, Germany and Ukraine. To allow further analysis and future research on the topic of pedestrian behavior, we released our dataset in the form of 346 short video clips (on average 5-15s). The data is annotated with ground truth information including bounding boxes for pedestrian detection, behavioral data of pedestrians and the driver of the recording vehicle, and contextual information such as demographics (e.g. gender and age of pedestrians), ambient conditions (e.g.
weather, time of day) and environmental factors (e.g. signal, road structure).

An example of behavioral annotation is depicted in Fig. 4. We divide the behavioral data into two groups: pedestrian and driver behavior. For pedestrians there is a crossing label indicating the period when the crossing takes place. A crossing event starts when the pedestrian steps off the curb into the crosswalk and finishes when the pedestrian clears the crosswalk. The crossing label can co-occur with any other label. Labels such as ‘looking’ or ‘glancing’ reflect the attentive state of the pedestrian, showing at what points the pedestrian is looking towards the recording vehicle. Looking can also happen at the same time with other actions. For instance, a pedestrian can be crossing while looking at the traffic.

In addition, there are two types of labels, which describe the dynamic state (‘moving slow’, ‘moving fast’ or ‘standing’) and the reactions of the pedestrians. The speed of pedestrians are estimated qualitatively based on the judgment of the labelers. The reactions are either explicit, such as ‘nod’ and ‘handwave’, or implicit, such as ‘slow down’ and ‘speed up’. The explicit reactions can co-occur with any dynamic state. However, if the reaction is implicit, the dynamic state and reaction labels are mutually exclusive. The reason behind this exclusivity is that implicit reaction might change the dynamic state of the pedestrian. Therefore, splitting the pedestrian’s dynamic state better highlights how and when the transition has taken place. Consider the following scenario as an example: a pedestrian initial state is ‘moving slow’, he/she reacts by speeding up and his/her new dynamic state is ‘moving fast’.

The driver behavioral tags only contain the last two labels described above, one showing the state of the vehicle (e.g. ‘moving slow’ or ‘moving fast’) and the other the reaction of the driver (e.g. ‘slow down’ or ‘speed up’). The current state of the vehicle (moving slow or fast) is based on the actual speed of the vehicle at the time. Moving slow means the vehicle is driving below 20 km/h whereas moving fast means the vehicle is moving with speed equal to 20 km/h or higher. The changes in the speed reflect the reaction of the driver whether he is slowing down, stopping or speeding up. It should be noted that similar to pedestrians, the driver behavioral labels are mutually exclusive.

In addition, our dataset has temporal correspondences between the frames meaning that each pedestrian has a unique id throughout the sequence and can be tracked easily.

Attempts have been made to increase the variability of the data to capture more diverse pedestrian behaviors. Our data is collected in different seasons and contains samples with different weather (snowy, rainy or sunny) and illumination (different time of the day) conditions (see Fig. 5).

The data also contains various road structures including urban streets, regional roads and parking lots. Each of these structures has different characteristics in terms of availability of traffic signals, density of the crowd or driving speed limits. Such characteristics may give rise to different behavioral patterns, which are important for our analysis.

Our proposed dataset is called Joint Attention in Autonomous Driving (JAAD) 3.

C. Behavioral labeling procedure

There are two main challenges when it comes to labeling behavioral data: uncertainty about the intention of the study subjects and subjective bias of the labeler in assessing the data. We tried to minimize these effects by using multiple labelers to process our data. If the labelers could not achieve consensus about the nature of a pedestrian action, it was excluded from the study to minimize error.

D. Pedestrian Samples

We identified more than 2600 pedestrians in our data, out of which we annotated 654 pedestrians crossing or near crossing (the pedestrian intends to cross but does not do so for some reasons). Unfortunately, not all of these samples contain a full crossing event, i.e. showing the pedestrian before, during and after the crossing. Also, in some cases the pedestrians intention was ambiguous. Therefore, we excluded 208 samples from the analysis, leaving a total of 446 samples.

For the analysis we only looked at the pedestrians’ age. Thus we categorized pedestrians as children (mid teen and below), adults and the elderly (above 65). Other factors such as gender and group size were excluded from the study.

3The JAAD dataset can be obtained at http://data.nvision2.eecs.yorku.ca/JAAD_dataset/. Ethics certificate # 2016-203 from York University.
IV. OBSERVATIONS AND ANALYSIS

A. Crossing patterns

We observed a high variability of pedestrian behaviors before and after crossing events as well as in the cases when the crossing did not take place. To quantify these behaviors for further analysis we collected action labels for each pedestrian and sorted them by the start time. This led to a list of more than a hundred unique sequences of actions. We visualized these sequences in Fig. 6 for crossing and no-crossing events separately.

In the figure, the beginning of each new action is marked with a vertical bar and curved lines are used to show connections between consecutive actions. The thickness of each line reflects the frequency of the action occurrence in the dataset. We categorize actions based on their semantic meaning into 5 types: action, precondition to crossing, attention, reaction to driver’s actions, crossing. These are shown in the diagram in different colors. Note that only some actions belong to a single type. For example, attention includes only ‘looking’ and ‘glancing’, whereas ‘standing’ may be either a precondition to crossing (e.g. standing at the curb) or a reaction to driver’s action (e.g. stopping when the driver did not yield). The diagram shows actions and overlapping actions, however, it demonstrates the variability of pedestrian behavior and uneven distribution of occurrences of certain actions. The driver’s actions are not explicitly shown in order to simplify the diagram.

The diagram in Fig. 6a shows 345 sequences of pedestrian’s actions prior to and after the crossing takes place. Two patterns, namely ‘standing, looking, crossing’ and ‘moving, looking, crossing’, describe more than 1/3 of the crossing events. This means that many pedestrians attend to traffic as they are waiting at the curb or approaching the road before crossing. The remaining two-thirds of the crossing scenarios are more complex and involve multiple actions before and after the crossing. For instance, a pedestrian may move towards the curb, stop, attend to the traffic, acknowledge the yielding driver by nodding and finally cross the street, while checking again whether it is safe to cross (‘moving, looking, nod, crossing, looking’). In rare ambiguous situations pedestrians and drivers may go through a cycle of actions and reactions before one of them yields to the other.

Similarly, in no-crossing events, 1/3 of all pedestrians are waiting at the curb and observing the traffic, which corresponds to the ‘standing, looking’ sequence in Fig. 6b. In other scenarios pedestrians may have started crossing already but are forced to clear the way (‘clear path’), slow down or stop if the driver is not giving them the right of way. Pedestrians also may yield to the drivers. In one of the videos, the pedestrian is approaching the road while looking at the traffic, slows down, waves his hand at the approaching car and stops to let it pass (‘moving, looking, slow down, handwave, standing, looking’).

The diagram also shows the frequency of occurrence of actions at certain points during the crossing/no-crossing events. This is reflected in the vertical bar height (taller bars correspond to more common actions) as well as the thickness of the lines connecting this bar to the next (frequency for different types of subsequent actions). For example, it can be inferred that most no-crossing events observed in the data start with pedestrians standing at the curb. In approximately 20% of the crossing events, pedestrians, who already started crossing, look again at the traffic to check that it is safe to continue.

B. Non-verbal communication

In more than 90% of the clips in our dataset, we observed some form of non-verbal communication. Perhaps the most prominent form of body language (which was present in all
these cases) is the change in pedestrians’ head orientation. Head movement and looking towards the approaching traffic is often a strong indicator of pedestrian crossing intention. In fact, out of the total number of instances of pedestrians’ head movements, more than 80% occurred prior to crossing. The remaining 20% were the cases when the vehicles were fully stopped behind a traffic signal.

Changes in head orientation also serve as evidence of pedestrians paying attention to their surroundings. For example, they may check the state of the traffic signal or the distance to the approaching vehicle. This assessment can happen either before or during crossing. Sometimes pedestrians do not show initial intention of crossing but might attend to the traffic while crossing (which accounted for 20% of the attention occurrences).

Pedestrian head orientation can be either in the form of looking (90% of the time) or glancing (10% of the time). Looking lasts for a longer period of time (more than 1s), whereas glancing is very brief (less than 1s) and is in the form of a quick head movement towards the traffic. The range of motion that pedestrians exhibit while paying attention also varies significantly. As illustrated in Fig. 7, head orientation can be very subtle (Fig. 7a) or rather extreme, involving major changes in body posture (Fig. 7d).

Other forms of communication are rarer and often are used as a sign of acknowledgment or response to the driver’s behavior. These non-verbal cues are either explicit or implicit. The explicit forms include nodding and hand gesture which could convey a different meaning depending on the context. For example, hand gesture is used as a form of showing gratitude, yielding or asking for the right of way. The implicit responses, on the other hand, are in the form of changes in pedestrian action including stopping, clearing path, slowing down or speeding up.

The occurrence frequency of attention and response behaviors is shown in Fig. 8.

C. Environmental factors

To analyze the effects of environmental factors, we consider the following two features that characterize a crosswalk: the presence of a traffic signal or a zebra crossing (Fig. 9), and the width of the street (Fig. 10). The former factor identifies a crosswalk as non-designated (no form of traffic signal, sign or zebra crossing is present), zebra crossing (either a zebra crossing or a pedestrian crossing sign is present), or signal (a traffic signal such as traffic light or stop sign is present). We treat the cases in which both a signal and a zebra crossing are available as signalized crosswalks.

Note that in the above categorization we differentiate between the crosswalks with only zebra crossing (or/and pedestrian crossing sign) and the ones with a signal. The reason is that traffic signals have a stronger prohibitive strength in forcing vehicles to yield to pedestrians. For example, a vehicle must stop before a stop sign or a red light, whereas it can chose whether to yield to the pedestrian when there is only zebra crossing present. Naturally, pedestrians are more cautious when crossing a street without a signal compared to crossing at a signalized crosswalk.

Being more cautious means that pedestrians are more likely to assess their surroundings prior to crossing. This was confirmed in our analysis. According to our data, pedestrians paid attention (turned their heads towards the traffic) in 81% of the times at non-designated crosswalks, compared to 69% at zebra crossings and only 36% at signalized crosswalks.

Besides attention frequency, the presence of traffic signals can be a direct indicator of how likely pedestrians will cross. To show this, we split our data into crossing and no-crossing...
events. Fig. 11a shows the distribution of the data in each category. By computing the likelihood of crossing for each crosswalk type (Fig. 11a) we can see that in over 95% of the cases pedestrians crossed the street (regardless of other factors) when some form of designated signal or zebra crossing was present. On the other hand, the crossing probability is less than 50% when there is no zebra crossing or traffic signal available (see Fig. 11b).

Street width also plays a role in defining pedestrian behavior. The types of streets in our data can be grouped into narrow (1 and 2 lanes) and wide (3 and 4 lanes) streets. Narrow streets are mainly residential roads in which vehicles drive slowly. Wide streets, on the other hand, are main roads where traffic is usually bidirectional and vehicles commute with a higher speed.

Given the above characteristics, the width of the street can impact crossing in two ways: first, the width of the street determines how long it would take a pedestrian to cross. Therefore, the pedestrian’s crossing affordance decreases as the street width increases. Second, the fact that vehicles drive faster in wider streets, means that the associated risk of crossing is also higher. These two factors directly affect pedestrian behavior. According to our findings, on average pedestrians pay attention to the approaching traffic prior to crossing 87% of the times when crossing narrow streets, whereas they do so over 95% of the times in wide streets. This means that pedestrians are generally more cautious when crossing wider streets.

D. Pedestrian factors

In this section, we analyze the effect of pedestrian demographics, in particular, age on crossing behavior. Here, we are interested to see how age influences the frequency of attention prior to crossing. We found that the older a pedestrian, the more conservatively he/she behaves, therefore he/she will be more likely to pay attention to the traffic. In fact, in our data, the frequency of attention was below 40% for children, 72% for adults and 76% for the elderly.

Another finding is that the attention duration of pedestrians may vary. On average, adults tend to look at the traffic for 1.32s, children for 1.43s and elderly for 1.45s.

E. Dynamic factors

In this section we examine pedestrian gap acceptance in terms of Time-To-Collision (TTC). Based on our data, pedestrians on average cross the street between the TTC of 3 to 7 seconds. TTC is also shown to influence how frequently pedestrians will pay attention to the approaching traffic prior to crossing. As depicted in Fig. 12, there are no instances of pedestrians attempting to cross without paying attention when the TTC is below 3s. Furthermore, it can be seen that the frequency of crossing without paying attention is correlated with TTC, i.e. the chance of crossing without attention rises as the TTC increases. This is expected because the farther (or slower) the vehicles are, the safer pedestrians feel to cross.

We further split our findings into two groups of non-designated and designated (with a signal or a zebra crossing) crosswalks to measure the influence of street structure on attention occurrence. The results are illustrated in Fig. 13. Here, once again, we can see that when some form of traffic signal or designated crossing is present, pedestrians feel more certain about crossing, and thus, tend to pay less attention prior to crossing. In the case of designated crosswalks, the results are somewhat complementary. Pedestrians tend to pay more attention between the TTC of 2 to 5s, whereas they do so less between the TTC of 6 to 9 seconds, and, of course, much less when the TTC is 15s or higher meaning that the car is far...
away or moving very slow. As for non-designated crosswalks, in general, pedestrians do pay attention prior to crossing, and they do more so when the TTC is low.

TTC in conjunction with the street’s structure also influences the probability of attention occurrence prior to crossing. As shown in Fig. 14, although attention occurrence is less likely in narrow streets, overall, the higher the TTC gets, the lower is the chance of pedestrians paying attention when crossing both wide and narrow streets.

We found that pedestrians behave differently at different TTC values. For example, as shown in Fig. 15, children and adults, on average, spend less time assessing the traffic prior to crossing, the higher the TTC gets. This is expected because the associated risk of crossing decreases when the vehicles are far away or their speed is low. In the case of the elderly, however, we observe that the attention duration increases steadily, reaching a maximum point somewhere between the TTC of 7 to 8s and suddenly plummets when approaching the TTC of 10s. At first, this result might seem counterintuitive but a closer look at the data reveals that these are the cases where the vehicle is at a far distance. Given that the elderly take more care in developing their decisions they require more time for observing the traffic.

**F. Driver action and pedestrian crossing**

<table>
<thead>
<tr>
<th>DA</th>
<th>Speeds</th>
<th>Slows down</th>
<th>Stops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speeds</td>
<td>0.962</td>
<td>0.013</td>
<td>0.025</td>
</tr>
<tr>
<td>Slows down</td>
<td>0.196</td>
<td>0.643</td>
<td>0.161</td>
</tr>
<tr>
<td>Stops</td>
<td>0.25</td>
<td>0.6</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Besides the contextual elements we discussed earlier, driver actions are of vital importance in determining pedestrian crossing behavior. To investigate their impact, we divide driver actions into three types: when the driver either maintains the current speed or speeds up (speeds), reduces the speed (slows down), or comes to a full stop (stops). In addition, we divide the scenarios into non-designated, zebra crossing and traffic signal (see Section IV-C for definitions).

Table I shows the distribution of each scenario according to driver action. As can be seen, when no signal or zebra crossing is present, in less than 20% of the cases a crossing event takes place even if the vehicle had maintained or increased its speed. In these cases, either the TTC is very high (average of 25.7s) or there is traffic congestion and the pedestrian anticipates that the car would stop shortly (see Fig. 16 for an example). More than 98% of no-crossing scenarios occurred when the driver did not slow down or stop for the pedestrian. There were cases, however, in which the driver yielded to the pedestrian and stopped but the pedestrian did not cross and gave the right of way to the driver by making a hand gesture (see Fig. 17 for an example).

Our data suggests that pedestrians are still fairly conservative despite the presence of zebra crossings (or pedestrian

---

Fig. 13. Pedestrian attention frequency based on TTC.

Fig. 14. Occurrence of attention based on TTC and the structure of the street.

Fig. 15. The relationship between pedestrian attention duration and TTC.

Fig. 16. The pedestrian is crossing the street regardless of the driver’s action because he anticipates that the vehicle will stop due to traffic congestion.

Fig. 17. The pedestrian is giving the right of way to the driver, despite the driver stopping.
crossing signs). In this case, only 25% of pedestrians crossed the street when the driver did not slow down or stop. At the signalized crosswalks, however, we observe a very different pedestrian behavior. In about 60% of the cases pedestrians cross the street even though the driver speeds. This means that, in this context, driver action is almost irrelevant because pedestrians expect the driver to comply with the law and stop at the signal. It should be noted that no-crossing events only occurred at signalized crosswalks when the pedestrian chose not to cross and yielded to the driver.

V. Conclusion

Communicating with pedestrians and understanding their intentions is a complex problem. Not only is the state of pedestrians, such as their pose, head orientation or gait, an indicator of crossing intention, but also the context in which they are observed can play an important role.

Pedestrians often use explicit means of communication such as handwave to resolve conflicts in traffic scenes, e.g. yielding to the driver, requesting the right of way, etc. The variability of these behaviors is high and they can convey very different meaning depending on the context.

We showed that the elements present in a scene can help to predict what a pedestrian is going to do next. Street properties such as width, the presence of zebra crossings or traffic signals can determine pedestrians level of confidence while crossing.

In addition, the driver’s dynamic state with respect to the pedestrians is important. Factors such as TTC, which reflects the speed and the position of the vehicle, should be considered.

Our findings also suggest that there is an interrelationship between contextual elements. For instance, although the majority of pedestrians tend to look at the traffic prior to crossing, they do so less when the street is narrow or when TTC is high. This is also true if the crosswalk is signalized because pedestrians feel safer and are, therefore, less cautious while crossing.

Understanding the context of a traffic scene is not limited to what we investigated in this work. There are other aspects that need to be further studied such as environmental conditions, e.g. weather or lighting, social conditions, e.g. group vs individual behavior, and demographics of the participants, e.g. the elderly vs children.

Furthermore, to better understand the nature of communication that takes place between drivers and pedestrians, it would be beneficial to collect data that contains both the motions of pedestrians and the driver (e.g. by recording the driver’s facial expressions).

One particular limitation of this study was that our analysis was only based on a naturalistic dataset. Therefore, some subjectivity bias was present in judging pedestrians’ intention or determining whether they are looking towards the recording vehicle or not. This issue can be resolved by conducting a survey from drivers and pedestrians after crossing, asking them about their true intention or whether they noticed the vehicle (or the pedestrian’s gaze) prior to crossing. We intend to investigate this in the future.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the NSERC Canadian Field Robotics Network (NCFRN), the Air Force Office for Scientific Research (USA), and the Canada Research Chairs Program through grants to JKT.

REFERENCES


