The Thirty-Third International
FLAIRS Conference (FLAIRS-33)

Real-Time Detection and Prediction of Relative Motion of Moving Objects in Autonomous Driving

Lalintha G. Polpitiya,[1] **Kamal Premaratne**[2]
[1]The MathWorks, Inc., Natick, Massachusetts, USA, lpolpiti@mathworks.com
[2]University of Miami, Coral Gables, Florida, USA, kamal@miami.edu

Abstract

Autonomous driving vehicles must have the ability to identify and predict behaviors of surrounding moving objects (e.g., other vehicles, cyclists, and pedestrians) in real-time. This is especially true in urban environments, where interactions become more complex due to high volumes of traffic. The work in this paper harnesses the Dempster-Shafer (DS) theoretic framework's ability to capture and account for various types of evidence uncertainty to develop a robust event detection and prediction model, which is appropriately calibrated to account for the underlying uncertainty so that it may be employed to arrive at a more informed decision.

Introduction

Advances in sensor and AI technologies have propelled rapid developments in autonomous driving during the past decade. Numerous research projects (DARPA 2014) aimed at solving the autonomous driving challenges have accelerated these developments. Today, scientists are diligently working at making fully self-driving cars a reality.

Fully driver-less cars, however, require effective methodologies for handling uncertainties and exceptions that may arise from the behavior of nearby vehicles (e.g., sudden turns at intersections and lane changes, mechanical breakdowns, vehicles blocking the roadway, etc.), cyclists and pedestrians (e.g., abrupt movement onto the road), navigating in highly congested traffic situations, bad weather, and other unpredictable and unanticipated situations. With the increased automation required by the next levels of autonomous driving tasks as identified in the Society of Automotive Engineers (SAE) taxonomy (SAE 2018), there is heightened interest in how to manage these uncertain and exception conditions.

The lack of effective methods to handle such situations is considered perhaps the most serious roadblock preventing the development of fully autonomous vehicles (AVs) (Ottley 2019). Current exceptions and uncertainty handling mechanisms rely heavily on the involvement of the human driver (Burns 2018; Kaber 2018). Irrespective of whether a human is involved in the decision-making process, we believe that the information provided should be properly calibrated to capture the underlying uncertainty so that the decision-maker can make a fully informed decision. The DS theoretic (DST) (Dempster 1968; Shafer 1976) models are ideally suited to capture and propagate these underlying uncertainties throughout the decision-making pipeline.

We examine this supposition by developing a DST model to capture the intended relative movements of other vehicles and pedestrians. With the predicted behavior of a pedestrian crossing the road or a vehicle making a turn, together with the uncertainty associated with this prediction in hand, the AV (or the human-in-the-loop) can arrive at a more informed decision as to the action it could take. In turn, with an appropriately calibrated model which captures the underlying uncertainty as the basis on which better more informed decisions are made, this work has the potential to significantly contribute to reduce driver fatigue and the number of traffic accidents, solve the problem of traffic congestion, allow individuals who do not (or cannot) drive more freedom, as well as make commuting easier and safer.

This paper is organized as follows: First, we provide key terminology in common use for autonomous driving and DST basic notions. This is followed by the proposed model for lateral and longitudinal relative motion of surrounding moving objects. The experimental results are provided next, followed by concluding remarks.

Preliminaries

Levels in Autonomous Driving. The level of driver attention, and how much driver action is needed to perform the driving task, are important factors to consider when attempting to define the level of autonomy. The SAE defines six levels of driving automation (SAE 2018): 0 (no automation), 1 (driver assistance), 2 (partial automation), 3 (conditional automation), 4 (high automation), and 5 (full automation). These levels depend on whether the vehicle can perform all or part of the *dynamic driving task* (DDT). In levels 1-2, the human driver monitors the driving environment; in levels 3-5, an *automated driving system* (ADS) performs all or part of the DDT on a sustained basis.

The Driving Task. The driving task that an AV would have to execute involves the following (SAE 2018): *(a) Lateral control:* Turning left, right, tracking a turn. *(b) Longi-*

tudinal control: Maintaining set speed and appropriate gap, detecting preceding vehicle, appropriate braking (acceleration/deceleration to maintain speed). *(c) Object and Event detection:* Real-time environmental object and event detection, recognition, classification and response preparation, i.e., first to identify the object (e.g: car, cyclist, pedestrian etc), and second to understand the motion of the object and see if we can infer what it will do next. *(d) Planning:* Informed decision-making aspect which follows the information gathering via perception and behavior prediction. This includes long and short term plans needed to travel to a destination, maneuvers such as lane changes, intersection crossings, and driving around other moving objects or traffic participants. *(e) Miscellaneous:* Identifying actions such as hand-waving, signaling, and interacting with drivers.

DST Basic Notions. In DS theory, the *frame of discernment (FoD)* refers to the set of all possible mutually exclusive and exhaustive propositions (Shafer 1976). We consider the case where the FoD is finite and we denote it as $\Theta = \{\theta_0, \theta_1, \ldots, \theta_{n-1}\}$. Proposition $\{\theta_i\}$, which is referred to as a *singleton,* represents the lowest level of discernible information. The power set of Θ, denoted by 2^Θ, form all the propositions of interest in DS theory. A proposition that is not a singleton is referred to as a *composite.* The set $A \backslash B$ denotes all singletons in $A \subseteq \Theta$ that are not included in $B \subseteq \Theta$, i.e., $A \backslash B = \{\theta_i \in \Theta \mid \theta_i \in A, \theta_i \notin B\}$. We use $\overline{A}$ to denote $\Theta \backslash A$ and $|A|$ to denote the cardinality of A.

Definition 1. *The mapping* $m : 2^\Theta \mapsto [0, 1]$ *is said to be a* basic belief assignment (BBA) *or a* mass assignment *if* $m(\emptyset) = 0$ *and* $\sum_{A \subseteq \Theta} m(A) = 1$. *The* belief *assigned to* $A \subseteq \Theta$ *is* $Bl : 2^\Theta \mapsto [0, 1]$ *where* $Bl(A) = \sum_{B \subseteq A} m(B)$. ■

So, the mass captures the 'support' that is strictly allocated to a given proposition and the belief assigned to a proposition takes into account the support for *all* of its subsets. The mass of a composite proposition (a general focal element) is free to move into its subsets (e.g., into individual singletons), which allows one to model the notion of *ignorance.* Complete ignorance is captured via the *vacuous BBA* 1_Θ*:* $m(A) = 1$ for $A = \Theta$, and $m(A) = 0$ for $A \subset \Theta$. Propositions that possess non-zero mass are referred to as *focal elements;* the set of all focal elements in an FoD is referred to as its *core* $\mathfrak{F}$, i.e., $\mathfrak{F} = \{A \subseteq \Theta \mid m(A) > 0\}$. Note that $|\mathfrak{F}|$ is the number of focal elements. $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ is referred to as the *body of evidence (BoE).*

Relative Motion of Adjacent Moving Objects

DST Model. We now develop a DST model which captures the uncertainties that are inherent in the evidence generated by vehicles, pedestrians, cyclists, and other moving objects in uncertain road situations. To this end, we consider the *lateral relative movement* (as *Left, Center,* or *Right*) and the *longitudinal relative movement* (as *Away, Stationary,* and *Toward*) relative to the AV. Fig. 1 illustrates these movements when another car, a cyclist, and a pedestrian are present near the AV. These categories will be further divided into *slow* and *fast,* giving ten categories of movement as *Fast Left (FL), Slow Left (SL), Center (C), Slow Right (SR),* and *Fast Right (FR)* for lateral movement and *Fast Away (FA), Slow Away (SA), Stationary (S), Slow Toward (ST),* and *Fast Toward (FT)* for longitudinal movement.

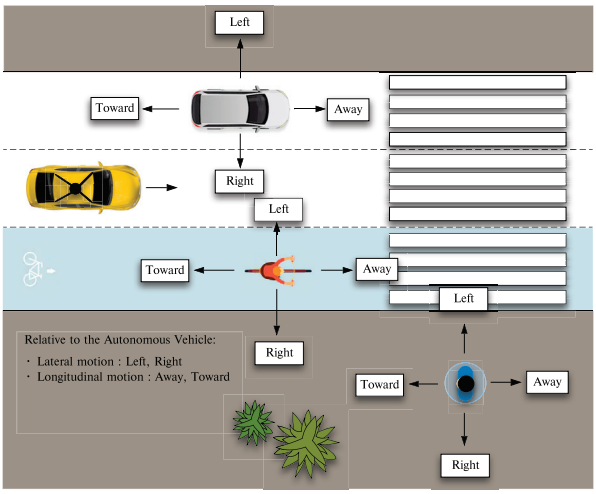

Figure 1: Lateral and longitudinal relative motion of moving objects. Lateral motion: *Left, Center,* and *Right;* longitudinal motion: *Away, Stationary,* and *Toward.*

Each moving object in the near vicinity of the AV is allocated a pair of lateral and longitudinal DST BoEs with the following two DST FoDs:

$$\begin{aligned} &\text{Lateral FoD:} && \Theta_U = \{FL, SL, C, SR, FR\}; \\ &\text{Longitudinal FoD:} && \Theta_V = \{FA, SA, S, ST, FT\}. \end{aligned} \quad (1)$$

Initialization of the BoEs. Let us link track T_i which may span several frames of a video signal, with object O_i. For n objects of interest, we have the tracks $\{T_0, T_1, ..., T_{n-1}\}$. We denote the object O_i in the previous and current frames as O_i' and O_i'', respectively. We denote the centroid of O_i' as $\{x', y'\}$, and the width and height of the relevant bounding box as w' and h', respectively. Similarly, the centroid of O_i'' is $\{x'', y''\}$, and the width and height of the relevant bounding box are w'' and h'', respectively. See Fig. 2.

Consider the object O_i on track T_i. For convenience, we will ignore the subscript i. The objects O' and O'', the object O (again, we ignore the subscript i) in the previous and current frames respectively, are endowed with two BoEs, one defined on the lateral FoD Θ_U and the other defined on the longitudinal FoD Θ_V (see (1)). Let us denote these BoEs as $\mathcal{E}_U' = \{\Theta_U, \mathfrak{F}_U', m_U'(\cdot)\}$ and $\mathcal{E}_V' = \{\Theta_V, \mathfrak{F}_V', m_V'(\cdot)\}$ for the previous frame and $\mathcal{E}_U'' = \{\Theta_U, \mathfrak{F}_U'', m_U''(\cdot)\}$ and $\mathcal{E}_V'' = \{\Theta_V, \mathfrak{F}_V'', m_V''(\cdot)\}$ for the current frame.

The starting BoE (the BoE associated with the first frame of track T_i) is taken to be vacuous, i.e.,

$$m(A) = 1_{\Theta_U}(\text{or } 1_{\Theta_V}) \equiv \begin{cases} 1, & \text{for } A = \Theta_U \text{ (or } \Theta_V); \\ 0, & \text{for } A \subset \Theta_U \text{ (or } \Theta_V). \end{cases} \quad (2)$$

Thus, the belief and plausibility of each proposition (except the full FoD itself) is 0 and 1, respectively.

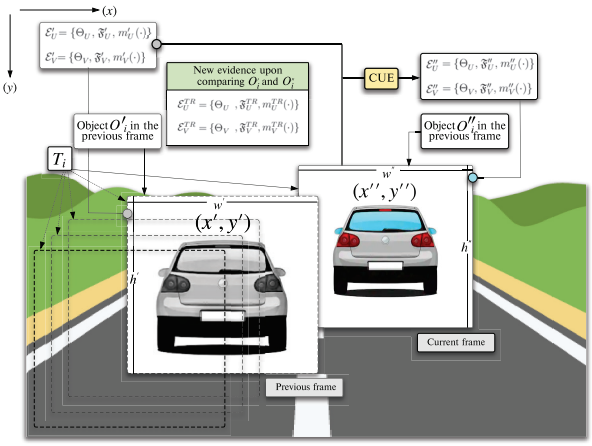

Figure 2: Updating the lateral and longitudinal BoEs of object O_i with CUE. $\mathcal{E}'' = (\mathcal{E}' \triangleleft \mathcal{E}^{TR})$. Previous frame: centroid $\{x', y'\}$, width w', height h'; current frame: centroid $\{x'', y''\}$, width w'', height h''.

DST Modeling of New Evidence. New evidence is extracted by comparing the relative lateral and longitudinal movement between O' and O''. Let us refer to these BoEs as *lateral and longitudinal transmit-BoEs,* and identify them as $\mathcal{E}_U^{TR} = \{\Theta_U, \mathfrak{F}_U^{TR}, m_U^{TR}(\cdot)\}$ and $\mathcal{E}_V^{TR} = \{\Theta_V, \mathfrak{F}_V^{TR}, m_V^{TR}(\cdot)\}$, respectively.

The BBAs associated with $\mathcal{E}_U^{TR}$ and $\mathcal{E}_V^{TR}$ are determined as follows. When an object is moving left or right, the value of x will change; when an object is moving away or toward, the value of y and also the area of the detected object will change. We use these changes to allocate DST masses which quantify the support the new evidence has for each proposition from the two FoDs Θ_U and Θ_V.

Algorithm 1 shows how these DST masses are computed. We use two constant parameters in this algorithm: Π, Γ: Π is used to classify the lateral focal element as either *FL* or *SL* (or either *FR* or *SR*); and Γ is used to classify the longitudinal focal element as either *FA* or *SA* (or either *FT* or *ST*). The exact values of Π and Γ depend on the configurations of the video input. Another parameter used in Algorithm 1 is the *detection confidence score* S which captures the confidence of the new evidence generated from O_i'' via the AV's pedestrian or vehicle detector.

The steps of Algorithm 1 are as follows: **(a)** *Line #1* takes as inputs $x', y', x'', y'', BBA_U^{TR}, BBA_V^{TR}, \Pi, \Gamma, S$. The centroid of O_i' is $\{x', y'\}$ and the centroid of O_i'' is $\{x'', y''\}$. We represent the BoEs BBA_U^{TR} and BBA_V^{TR} as *DS-Vectors* (although one could also employ *DS-Matrix* or *DS-Tree* structures) (Polpitiya et al. 2016; 2017). Π and Γ values are constants for the entire run; the value S varies with every new detection. **(b)** *Lines #2–12* determine the lateral focal element and assigning the mass. Mass of the lateral focal element will be S. **(c)** *Line #13* assigns value $1 - S$ to $m_U^{TR}(\Theta_U)$. **(d)** *Lines #14–24* determine the longitudinal focal element; its mass is S. **(e)** *Line #25* assigns value $1 - S$ to $m_V^{TR}(\Theta_V)$. **(f)** *Line #26* returns BBA_U^{TR} and BBA_V^{TR}

Algorithm 1 Determining the masses of $\mathcal{E}_U^{TR}$ and $\mathcal{E}_V^{TR}$

```
1: procedure AssignMasses(x', y', x'', y'', BBA_U^TR, BBA_V^TR, Π, Γ, S)
2:     if x' > Π + x'' then
3:         m_U^TR(FL) ← S
4:     else if x' > x'' then
5:         m_U^TR(SL) ← S
6:     else if x' + Π < x'' then
7:         m_U^TR(FR) ← S
8:     else if x' < x'' then
9:         m_U^TR(SR) ← S
10:    else
11:        m_U^TR(C) ← S
12:    end if
13:    m_U^TR(Θ_U) ← 1 − S
14:    if y' > Γ + y'' then
15:        m_V^TR(FA) ← S
16:    else if y' > y'' then
17:        m_V^TR(SA) ← S
18:    else if y' + Γ < y'' then
19:        m_V^TR(FT) ← S
20:    else if y' < y'' then
21:        m_V^TR(ST) ← S
22:    else
23:        m_V^TR(S) ← S
24:    end if
25:    m_V^TR(Θ_V) ← 1 − S
26:    Return BBA_U^TR, BBA_V^TR
27: end procedure
```

DS-Vectors.

Evidence Updating. For updating the BoEs, we use the conditional update equation (CUE) and conditional fusion equation (CFE):

Definition 2 (CUE)**.** *(Premaratne et al. 2009) Suppose the receive-BoE* $\mathcal{E}' = \{\Theta, \mathfrak{F}', m'(\cdot)\}$ *is presented with the evidence provided by the transmit-BoE* $\mathcal{E}^{TR} = \{\Theta, \mathfrak{F}^{TR}, m^{TR}(\cdot)\}$*. Then, the* CUE-based BBA update *of the receive-BoE* $\mathcal{E}''$ *is*

$$m''(B) = \alpha\, m'(B) + (1-\alpha) \sum_{A \in \mathfrak{F}^{TR}} \beta(A)\, m^{TR}(B|A),$$

for all $B \subseteq \Theta$*. Here, the CUE parameters* $\{\alpha, \beta(\cdot)\}$ *are non-negative real and satisfy*

$$\alpha + (1-\alpha) \sum_{A \in \mathfrak{F}^{TR}} \beta(A) = 1,\ \forall k \in \mathbb{N}_+.$$

We denote this operation as $\mathcal{E}'' = (\mathcal{E}' \triangleleft \mathcal{E}^{TR})$. ■

Definition 3 (CFE)**.** *(Wickramarathne, Premaratne, and Murthi 2012) The* CFE *that fuses the evidence of* $\mathcal{E}^{TR1} = \{\Theta, \mathfrak{F}^{TR1}, m^{TR1}(\cdot)\}$ *and* $\mathcal{E}^{TR2} = \{\Theta, \mathfrak{F}^{TR2}, m^{TR2}(\cdot)\}$ *is*

$$m^{TR}(B) = K^{TR1} \sum_{A \in \mathfrak{F}^{TR1}} \beta^{TR1}(A)\, m^{TR1}(B|A) + K^{TR2} \sum_{A \in \mathfrak{F}^{TR2}} \beta^{TR2}(A)\, m^{TR2}(B|A),$$

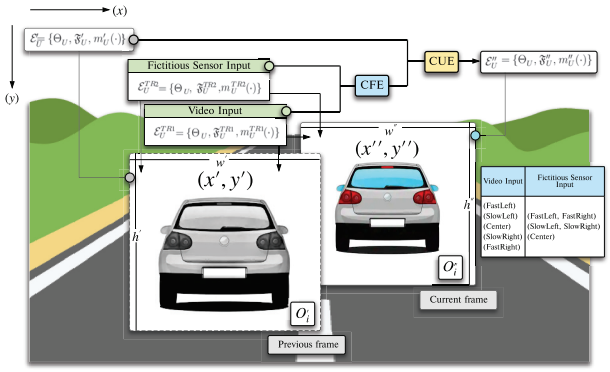

Figure 3: Updating the lateral and longitudinal BoEs of object O_i. $\mathcal{E}^{TR} = (\mathcal{E}^{TR1} \bowtie \mathcal{E}^{TR2})$, $\mathcal{E}'' = (\mathcal{E}' \triangleleft \mathcal{E}^{TR})$.

for all $B \subseteq \Theta$. *The parameters* $\{K^{TR1}, K^{TR2}, \beta(\cdot)\}$ *are non-negative real and satisfy*

$$K^{TR1} \sum_{A \in \mathfrak{F}^{TR1}} \beta^{TR1}(A) + K^{TR2} \sum_{A \in \mathfrak{F}^{TR2}} \beta^{TR2}(A) = 1.$$

We denote this operation as $\mathcal{E}^{TR} = (\mathcal{E}^{TR1} \bowtie \mathcal{E}^{TR2})$. ■

The CUE is used to update a given BoE when it confronts a new incoming BoE (see Experiment 1 below). Fig. 2, illustrates the CUE update operation, $\mathcal{E}'' = (\mathcal{E}' \triangleleft \mathcal{E}^{TR})$. We determine the CUE parameters $\{\alpha, \beta(\cdot)\}$ via the following strategy: **(a)** *Selecting* α*:* We use a constant value for α so that the incoming evidence is given the same weight throughout the update process. The value of this constant α is a function of the fidelity of the input (e.g., the resolution and frame rate of the video input) and the particular domain of application. **(b)** *Selecting* $\beta(\cdot)$*:* These are selected according to the *receptive update strategy* (Premaratne et al. 2009), i.e., $\beta(A) = m^{TR}(A),\ \forall A \in \mathfrak{F}^{TR}$. In essence, receptive updating 'weighs' the incoming evidence according to the support each focal element receives from $\mathcal{E}^{TR}$. In the special case of probability mass functions, this receptive strategy reduces to a weighted average of probability mass functions corresponding to $\mathcal{E}'$ and $\mathcal{E}^{TR}$.

The CFE is used to generate a new BoE by fusing two (or more) incoming BoEs (see Experiment 2 below). Fig. 3, illustrates the CFE fusion operation, $\mathcal{E}^{TR} = (\mathcal{E}^{TR1} \bowtie \mathcal{E}^{TR2})$. We determine the CFE parameters $\{K^{TR1}, K^{TR2}, \beta(\cdot)\}$ via the following strategy: (a) *Selecting* K^{TR1} *and* K^{TR2}*:* We use constant values for K^{TR1} and K^{TR2} so that the fusing evidences are given the same weight throughout the fusion process. The values of these constants K^{TR1} and K^{TR2} are a function of the properties of the sensor inputs. **(b)** *Selecting* $\beta(\cdot)$*:* These are also selected according to the *receptive update strategy*, i.e., $\beta(A) = m^{TR1}(A),\ \forall A \in \mathfrak{F}^{TR1}$. The receptive updating 'weighs' the incoming evidence according to the support each focal element receives from the BoEs $\mathcal{E}^{TR1}$ and $\mathcal{E}^{TR2}$.

Experiments

We conducted experiments of these models of evidence and updating on two AV navigation tasks: pedestrian navigation and vehicle lane change. These experiments were carried out with the help of the following `MATLAB` toolboxes: Automated Driving, Computer Vision, and Deep Learning (https://www.mathworks.com/products/automated-driving.html, computer-vision.html, deep-learning.html).

Experimental Setup.

Video Input. We used a 30 frames/s (FPS) video input with (640×480)-pixel (480p) resolution. The video input was obtained using a Canon EOS Rebel T3i Digital Single Lens Reflex (DSLR) camera with a EF-S 18-55mm f/3.5-5.6 IS STM standard zooming lens. Following are the other camera intrinsic parameters: Focal length: [2.3814e+03, 1.8671e+03], Principal point: [283.7174, 282.2949], and Radial distortion: [7.1048, -95.8961]. Video inputs were collected in urban environments and highways in Florida during the day time.

Object Identification and Detection. We used the following pre-trained detectors: **(a)** For detection of vehicles, `vehicleDetectorACF('full-view')` and `vehicleDetectorFasterRCNN('full-view')`. from the Automated Driving Toolbox and the Deep Learning Toolbox. These 'full-view' models have been trained with images that contain unoccluded views from the front, rear, left, and right sides of vehicles. **(b)** For detection of pedestrians and cyclists, `peopleDetectorACF('caltech')` from the Computer Vision Toolbox. It has been trained using the Caltech Pedestrian data set which contains approximately 10 hours of 640x480 30Hz video taken from a vehicle driving through regular traffic in an urban environment and includes 350,000 pedestrian bounding boxes labeled in 250,000 frames (Dollar et al. 2012).

Object Tracking. We implemented a multiple object tracker using the Computer Vision Toolbox. It keeps an array of tracks, where each track is a structure representing a pedestrian or a vehicle in the video. The purpose of the structure is to maintain the state of a tracked moving object. The tracker estimates the state vector and state vector covariance matrix for each track using a Kalman filter. These state vectors are used to predict a track's location in new frames and to calculate the likelihood of each detection being assigned to each track.

Modeling and Updating of Evidence. Both DST and probability models were employed to capture evidence. We updated the belief, plausibility, and probability values for each object as each frame was received.

DST Model. The FoDs associated with the lateral and longitudinal movements are Θ_U and Θ_V, respectively. With the initial BoEs taken as vacuous, each BoE was updated using the CUE, for which after calibration and a series of experiments, we used the parameter value $\alpha = 0.66$; the receptive update strategy was employed to select the β values.

The fusion operations were carried out with the CFE, for which we used the parameter values, $K^{TR1} = 0.5$ and $K^{TR2} = 0.5$; the receptive update strategy was utilized to select the β values. Data structures, algorithms, DS-Conditional-One model and DS-Conditional-All model present in (Polpitiya et al. 2016; 2017; Polpitiya 2019) were utilized for DST implementations.

Probability Model. The sample spaces used were the

Figure 4: One frame of a video stream corresponding to a pedestrian crossing the road from left to right.

same, viz., Θ_U and Θ_V. Initially, having no evidence to use as a basis, we assumed that the probabilities are uniformly distributed, i.e., a probability of $1.0/5 = 0.2$ was used for each singleton. Probability updating was carried out in the following manner: Consider the object O'_i in the frame and its corresponding object O''_i in the current frame. Suppose the lateral and longitudinal probability distributions associated with O'_i and O''_i are $P'_U(\cdot)$, $P'_V(\cdot)$ and $P''_U(\cdot)$, $P''_V(\cdot)$, respectively. We get the lateral and longitudinal new probability distributions, which we receive as new evidence upon comparing the relative positions of O'_i and O''_i. Suppose these new probability distributions are $P^{TR}_U(\cdot)$ and $P^{TR}_V(\cdot)$. Using an update mechanism analogous to CUE, we generate the masses of $P^{TR}_U(\cdot)$ and $P^{TR}_V(\cdot)$ via

$$P''(\cdot) = \alpha\, P'(\cdot) + (1-\alpha)P^{TR}(\cdot), \text{ with } \alpha = 0.66. \quad (3)$$

We analyzed several scenarios under different driving and road conditions. Results of the scenarios being presented are the following: **(a)** Experiment 1: Pedestrian road crossing; **(c)** Experiment 2: Pedestrian road crossing with an additional fictitious sensor input incorporated to the video input.

Experiment 1

Fig. 4 shows one frame of the analyzed video of a pedestrian, starting from the middle of the road, crossing the road from left to right. The vehicle, whose passenger was videoing, was turning left inside a roundabout.

Fig. 5 shows the quantities $Bl()$, $Pl()$, and $P()$ versus the frame number. Consider the 4 segments ①-④.

① $Bl()$ values start from 0, $Pl()$ values start from 1, and $P()$ values are in between. With incoming evidence, the uncertainty intervals $Un() \equiv Pl() - Bl()$ are narrowed. Note how the DST model allows us to start with total ignorance (i.e., $Un() = 1$) instead of a point probability.

② Evidence received being of high confidence, $Bl()$, $Pl()$, and $P()$ values converge. The pedestrian motion prediction is SL (which is the ground-truth).

③ Here we receive evidence with low confidence, and while $Bl(SR)$ and $P(SR)$ values stay equal, the $Pl(SR)$ value is at 1. The ground-truth is contained within $Un(SR)$. The $P(SR)$ value gets lower and moves away from the ground-truth; the other $P()$ values are increasing.

④ At the end of the object detection, $Un()$ increases while $Bl(SR)$ and $P(SR)$ values stay equal. Eventually $P(SR)$ reaches a near zero value which contradicts the ground-truth (the pedestrian is moving to the right and out of the frame). While the other $P()$ values are also increasing, $Bl()$ and $Pl()$ values correctly represent the uncertainty.

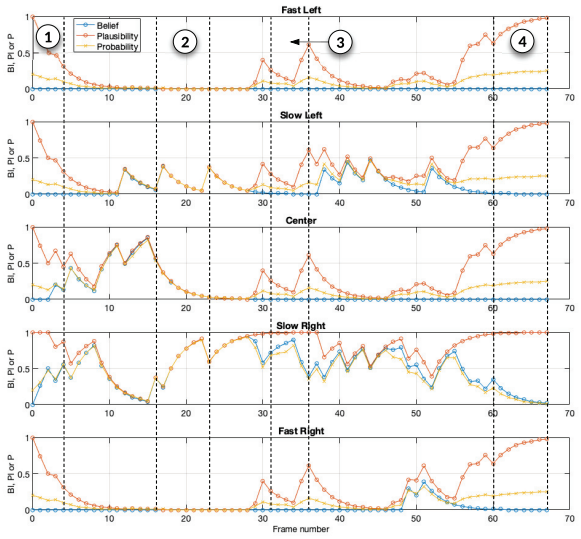

Figure 5: $Bl()$, $Pl()$ and $P()$ values when a pedestrian is crossing the road from left to right.

Experiment 2

This experiment uses the evidence from the same video stream utilized above (denoted by the BoE $\mathcal{E}^{TR1}$) together with the evidence from an extra fictitious sensor input (denoted by the BoE $\mathcal{E}^{TR2}$). The CFE is first applied to fuse $\mathcal{E}^{TR1}$ and $\mathcal{E}^{TR2}$; then the CUE is applied as before the update the evidence (see Fig. 3).

The BoEs for the two sensor streams are generated as follows: **(a)** $\mathcal{E}^{TR1}$*:* Here we apply Algorithm 1. The constant value Π classifies the lateral focal element as either *FL* or *SL* (or either *FR* or *SR*) and it depends on the video input configuration. The constant Γ classifies the longitudinal focal element. **(b)** $\mathcal{E}^{TR2}$*:* Here we employ an algorithm similar to Algorithm 1. The constant Π is used to discriminate the lateral focal element as either *Slow* or *Fast* (not as *FL*, *SL*, *C*, *FR*, *SR*). Focal elements are not discriminated as *Right* or *Left* when using the fictitious sensor. This ensures that the BoEs generated are not necessarily consonant.

Fig. 6 shows the quantities $Bl()$ and $Pl()$ versus the frame number. Results of Experiment 1 are also provided in the figure for comparison purposes. Note that the results are obtained in real-time. Each fusion and updating operation is completed within 33 (ms).

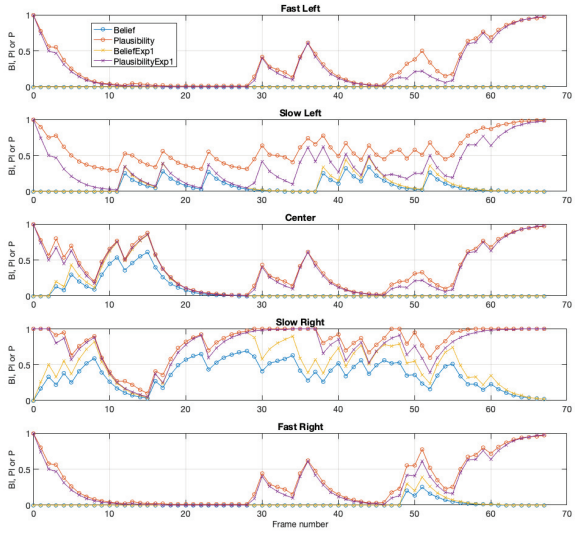

Figure 6: $Bl()$ and $Pl()$ values when a pedestrian is crossing the road from left to right. Simulation results with two sensor inputs.

Concluding Remarks

The absence of effective mechanisms to handle uncertain and exception conditions is perhaps the most serious hindrance toward the development of fully autonomous vehicles (Ottley 2019). The current approach of using deep learning methods to improve decision-making must rely on 'experience' captured via training data. However, most traffic accidents are due to exceptions (e.g., bad weather, traffic violations, sudden movement of other moving objects and pedestrians, mechanical breakdowns, etc.). Achieving full autonomy calls for appropriate mechanisms for handling data/evidence uncertainty and exceptions. Currently, exception handling relies on the human driver, and it is not practical to assume that a driver is always receptive to take over when an AV issues a *request to intervene*. Irrespective of whether it is the AV itself or a human driver who has to take control of the situation, the decision-maker will be better served if a fully informed decision which accounts for the underlying uncertainty can be made.

We explore the utility of DST models as a way to capture the uncertainty in intended relative movements of other vehicles and pedestrians. The calibrated information that such a model generates may allow an AV (or a human-in-the-loop) to make more informed decisions. To ensure real-time operation, we have made use of the data structures and algorithms in (Polpitiya et al. 2016; 2017; Polpitiya 2019).

We believe that the proposed model is a significant step toward addressing the current challenges involving levels 4-5 autonomous driving tasks. DST models are ideal to capture and propagate the uncertainties throughout the decision-making pipeline and provide an interval-valued measure of the uncertainty underlying any inference it makes.

Acknowledgments

This work is based on research supported by the U.S. Office of Naval Research (ONR) via grant #N00014-10-1-0140, U.S. National Science Foundation (NSF) via grant #1343430, and the University of Miami U-LINK Initiative on Interdisciplinary Inquiry.

References

Burns, C. M. 2018. Automation and the human factors race to catch up. *Journal of Cognitive Engineering and Decision Making* 12(1):83–85.

DARPA. 2014. The DARPA Grand Challenge: Ten Years Later. https://www.darpa.mil/news-events/2014-03-13.

Dempster, A. P. 1968. A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B* 30(2):205–247.

Dollar, P.; Wojek, C.; Schiele, B.; and Perona, P. 2012. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(4):743–761.

Kaber, D. B. 2018. Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. *Journal of Cognitive Engineering and Decision Making* 12(1):7–24.

Ottley, S. 2019. Nissan: Autonomous cars impossible without humans. https://www.drive.com.au/news/nissan-autonomous-cars-impossible-without-humans-120948.

Polpitiya, L. G.; Premaratne, K.; Murthi, M. N.; and Sarkar, D. 2016. A framework for efficient computation of belief theoretic operations. In *Proceedings of the Nineteenth International Conference on Information Fusion (FUSION)*, 1570–1577.

Polpitiya, L. G.; Premaratne, K.; Murthi, M. N.; and Sarkar, D. 2017. Efficient computation of belief theoretic conditionals. In *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA)*, 265–276.

Polpitiya, L. G. 2019. *A Framework for Efficient Implementation and Effective Visualization of Dempster-Shafer Belief Theoretic Computations for Reasoning Under Uncertainty*. Ph.D. Dissertation, University of Miami, Coral Gables, FL.

Premaratne, K.; Murthi, M.; Zhang, J.; Scheutz, M.; and Bauer, P. 2009. A Dempster-Shafer theoretic conditional approach to evidence updating for fusion of hard and soft data. In *Proceedings of the Twelfth International Conference on Information Fusion (FUSION)*, 2122–2129.

SAE. 2018. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Technical report, SAE International.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press.

Wickramarathne, T. L.; Premaratne, K.; and Murthi, M. N. 2012. Consensus-Based credibility estimation of soft evidence for robust data fusion. In *Proceedings of the Second International Conference on Belief Functions (BELIEF)*, 301–309.