Ontology-based framework for risk assessment in road scenes using videos

Mahmud Abdulla Mohammad, Ioannis Kaloskampis, Yulia Hicks, Rossitza Setchi

School of Engineering, Cardiff University, Queen’s Buildings, The Parade, Cardiff, UK

Abstract

Recent advances in autonomous vehicle technology pose an important problem of automatic risk assessment in road scenes. This article addresses the problem by proposing a novel ontology tool for assessment of risk in unpredictable road traffic environment, as it does not assume that the road users always obey the traffic rules. A framework for video-based assessment of the risk in a road scene encompassing the above ontology is also presented in the paper. The framework uses as input the video from a monocular video camera only, avoiding the need for additional sometimes expensive sensors. The key entities in the road scene (vehicles, pedestrians, environment objects etc.) are organised into an ontology which encodes their hierarchy, relations and interactions. The ontology tool infers the degree of risk in a given scene using as knowledge video-based features, related to the key entities. The evaluation of the proposed framework focuses on scenarios in which risk results from pedestrian behaviour. A dataset consisting of real-world videos illustrating pedestrian movement is built. Features related to the key entities in the road scene are extracted and fed to the ontology, which evaluates the degree of risk in the scene. The experimental results indicate that the proposed framework is capable of assessing risk resulting from pedestrian behaviour in various road scenes accurately.

Keywords: Pedestrian detection, Video segmentation, Ontology;

1. Introduction

Recent advances in autonomous vehicles have resulted in intelligent automobiles which sense the environment using a variety of sensors, such as GPS, radars and cameras. By processing the information acquired by these sensors, they are capable of determining the travel route [1] and identifying important scene objects, such as traffic signs [2] and obstacles [3]. An important aspect in the design of autonomous vehicles is safety assessment in a given road scene, which is the problem of determining the degree of risk in the scene given a number of sensor measurements.

E-mail address: MohammadMA, KaloskampisI, HicksYA, Setchi @cardiff.ac.uk
Recognition of important scene objects around the vehicle is crucial when assessing the risk in a given road scene [4]. However, object recognition does not provide sufficient information to evaluate the situation with respect to safety, as the behaviour of these objects is also important. Fig. 1 shows two scenes featuring the same objects. In Fig. 1a the pedestrian is on the road and therefore the situation is riskier than Fig. 1b, where the pedestrian is on the pavement and moving away from the road. At the same time, certain environment factors influence the risk assessment of the scene, such as visibility condition (fog, haze pollution and light), weather, traffic signs, road type and road quality [5]. It is therefore apparent that the assessment of risk in a road scene involves the processing of a plethora of information, arising from several entities. These entities interact with each other. In the example in Fig. 1 for instance, the interactions between the pedestrian, the road and the pavement influence the degree of risk. The goal of this work is to represent this information using a model which takes all entities and their interactions into account and enables reasoning, i.e. assessment of risk given sensor measurements relevant to the modelled entities.

In the past, several different methods were employed to solve this problem. Platho et al. [6] decomposed the task of traffic situation assessment into sets of entities, with each set affecting one road user. The entities in each set are linked using a Bayesian network. However, there are no direct interactions between different sets and thus this method may have problems propagating the effect of events from one set to another. Schamm and Zöllner [7] use a knowledge-based framework which takes into account interactions between entities to solve the above problem. Vacek et al. [8] tackle the same problem using case-based reasoning. Their model is capable of updating its knowledge base with newly encountered behaviours; however, the system’s stability may be compromised when fed with an excessive number of situations [6].

Ontologies have been used successfully in the past to model efficiently complex interactions between entities in road scene environments and represent a wide variety of behaviours without stability issues. Hülsen et al. [9] proposed an ontology-based situation description method for traffic intersections. Pollard et al. [5] presented an ontology for situation assessment for automated ground vehicles taking into account the vehicle perception, environmental conditions and the driver’s ability. Information regarding these parameters was acquired using several different sensors (cameras, GPS, laser range finder sensors etc.). The purpose of the study was to determine the level of automation of a vehicle. An ontology-based situational awareness framework was proposed by Armand et al. [10], which utilises contextual information to infer the behaviour of the perceived entities (e.g. vehicles, pedestrians). However, their frameworks assume that pedestrians and subject vehicles obey the traffic rules, which is not always the case in real world traffic environment.

It is also worthy to note that in the above frameworks the information regarding the perceived entities is acquired using several types of sensors simultaneously. Although the presence of multiple sensors offers rich information, due to high cost, complex installation and high computational load it is currently not close to becoming standard for vehicles. Certain sensors such as ultrasonic, radar, and laser may additionally suffer from interference problems [11].

In this article an ontology-based framework for assessing the degree of risk in a road scene is proposed. The framework is built around a novel ontology which encompasses the key entities in the road scene and encodes their hierarchy, relations and interactions. The novelty of the framework stems from its ability to interpret unpredictable road traffic, as it does not assume that the road users always obey the traffic rules.

The proposed framework uses as input the data from a monocular video camera only, capturing footage from the driver’s perspective in an efficient, cost-effective approach [12,13]. The existence of a visual sensor is gradually becoming standard for modern vehicles as an increasing number of vehicles is equipped with dashboard cameras.

The evaluation of the proposed framework focuses on scenarios in which risk results from pedestrian behaviour. To assess the performance of our framework, we introduce a dataset consisting of real-world videos illustrating pedestrian movement captured from the driver’s point of view. The experimental results indicate that the proposed framework is capable of assessing risk resulting from pedestrian behaviour in a variety of road scenes with high accuracy.

The remainder of the paper is organized as follows. In Section 2, we discuss the proposed risk assessment method. Experimental results are given in Section 3. Finally, conclusions and future work are discussed in Section 4.

2. Risk assessment method

As explained in the introduction, assessing the degree of risk in a road scene is more challenging when considering the more general problem of interpreting the unconstrained behaviour of entities in the scene. In this work we propose
Fig. 1: Illustrating risk situation.

Fig. 2: Ontology structure.
a novel ontology-based framework for assessing the degree of risk in a road scene. Our ontology is designed to cater for risk related to several factors, such as risk from objects (vehicles, pedestrians, cyclists etc.), environmental risk (weather and visibility condition) and road environmental risk (road quality, road traffic signs and road types).

2.1. Ontologies

In philosophy, the ontology is defined as an “account of existence”[14]. In computer engineering, the definition of ontology is the “specification of a conceptualization”[14]. In more detail, ontology is the hierarchical definition of the terms and relationship between them, which is the formal representation of knowledge that understandable by humans and computer [10]. An ontology-based framework consists of a terminological box (TBox) that includes concepts, role definitions and axioms, and an assertional box (ABox) that includes instances of concepts, roles among such instances [5,10].

2.2. The structure of proposed ontology

The proposed ontology is shown in Fig.2. It consists of three main classes which correspond to factors contributing to risk: risk from object, environmental risk and road environmental risk. In the next paragraph, we discuss each of these classes individually.

The risk factor classes consist of several levels of subclasses. The structure of the ontology is organised on the basis of the relations between subclasses and main classes. In the following paragraphs we discuss the structure of the risk factor classes:

1. Risk from object: The role of this class is to provide detailed information regarding the object attributes, so that the degree of risk can be assessed from the type and behavior of each object in the scene. This class contains object attributes: object speed (with four subclasses representing different speed levels), object motion direction...
(with two subclass of object direction), and object type. Also, the object type class consists of two subclasses: stationary object (with three subclasses representing different stationary objects) and moving object (with four subclasses representing different types of moving objects).

2. Environmental risk: The role of this class is to provide a detailed description of the environment. This class consists of two subclasses: weather conditions and visibility conditions. Weather conditions class contains six subclass, rain (with three subclasses for different types of rain), snow (with four subclasses for different types of snowing), cloud (with three subclasses for different types of cloudiness) and wind. Wind consists of two subclasses, wind speed (with three subclasses as a different degree of speed) and wind direction (with four subclasses as a different direction), sun and temperature. Visibility conditions consist of three subclasses, fog (with three subclasses for different degrees of haze pollution) and light (with three subclasses for different types of lighting).

3. Road environmental risk: The role of this class is to provide rich information on the road environment, on the basis of which the risk from this factor can be assessed. This class consists of three subclasses: road quality (with subclass road surface quality), road type (with four subclasses for different road types), road traffic signs (with subclass warning traffic signs).

Object property is the binary relation between two classes. Here, thirteen object properties are defined based on the necessity of the relations, namely highRisk, mediumRisk, lowRisk, noRisk, hasHighSpeed, hasMediumSpeed, hasLowSpeed, hasNoSpeed, hasAwayFrom, hasTowardThe, objectOnTheRoad, objectOnTheRoadEdge, and objectOnTheRoadSide.

In this structure, to assess the risk level of the RiskAssessment, only one of the properties among highRisk, mediumRisk, lowRisk, noRisk must be inferred. Again, only one of the speed properties hasHighSpeed, hasMediumSpeed, hasLowSpeed and hasNoSpeed must be inferred, and these properties specify speed type ObjectSpeed of the movingObject. The object motion direction property ObjectMotionDirection of the movingObject according to the Road observer, is inferred by hasAwayFrom and hasTowardThe. Finally, the stationary object StationaryObject’s interaction with the moving object movingObject is inferred on the basis of one of the properties objectOnTheRoad, objectOnTheRoadEdge, and objectOnTheRoadSide.

2.3. Assessment of pedestrian risk

In this section we define both the risk assessments, generated by pedestrian behaviour in the scene (see Fig.3), and inference rules. These rules are based on our knowledge and the information from the risk factor classes. The rules are formed in semantic web rule language (SWRL).

1. High risk: The situation involves a high level of danger. It is inferred according to the following rules:

   Pedestrian (?)p ∧ Road (?)r ∧ hasHighSpeed (?)p, ?s) ∧ objectOnTheRoad (?)p, ?r) → highRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasMediumSpeed (?)p, ?s) ∧ objectOnTheRoad (?)p, ?r) → highRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasLowSpeed (?)p, ?s) ∧ objectOnTheRoad (?)p, ?r) → highRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasNoSpeed (?)p, ?s) ∧ objectOnTheRoad (?)p, ?r) → highRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasHighSpeed (?)p, ?s) ∧ objectOnTheRoadEdge (?)p, ?re)
   ∧ hasTowardThe (?)p, ?r) → highRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasMediumSpeed (?)p, ?s) ∧ objectOnTheRoadEdge (?)p, ?re)
   ∧ hasTowardThe (?)p, ?r) → highRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasLowSpeed (?)p, ?s) ∧ objectOnTheRoadEdge (?)p, ?re)
   ∧ hasTowardThe (?)p, ?r) → highRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasNoSpeed (?)p, ?s) ∧ objectOnTheRoadEdge (?)p, ?re) → highRisk (?)p, ?a)

2. Medium risk: The situation involves a medium level of danger. It is inferred according to the following rules:

   Pedestrian (?)p ∧ Road (?)r ∧ hasHighSpeed (?)p, ?s) ∧ objectOnTheRoadEdge (?)p, ?re)
   ∧ hasAwayFromThe (?)p, ?r) → mediumRisk (?)p, ?a)
   Pedestrian (?)p ∧ Road (?)r ∧ hasMediumSpeed (?)p, ?s) ∧ objectOnTheRoadEdge (?)p, ?re)
∧ hasAwayFromThe (?p, ?r) → mediumRisk (?p, ?a)
Pedestrian (?p) ∧ Road (?r) ∧ hasLowSpeed (?p, ?s) ∧ objectOnTheRoadEdge (?p, ?re)
∧ hasAwayFromThe (?p, ?r) → mediumRisk (?p, ?a)
Pedestrian (?p) ∧ Road (?r) ∧ hasNoSpeed (?p, ?s) ∧ objectOnTheRoadEdge (?p, ?re) → mediumRisk (?p, ?a)
Pedestrian (?p) ∧ Road (?r) ∧ hasHighSpeed (?p, ?s) ∧ objectOnTheRoadSide (?p, ?rs)
∧ hasTowardThe (?p, ?r) → mediumRisk (?p, ?a)
Pedestrian (?p) ∧ Road (?r) ∧ hasMediumSpeed (?p, ?s) ∧ objectOnTheRoadSide (?p, ?rs)
∧ hasTowardThe (?p, ?r) → mediumRisk (?p, ?a)
Pedestrian (?p) ∧ Road (?r) ∧ hasLowSpeed (?p, ?s) ∧ objectOnTheRoadSide (?p, ?rs)
∧ hasTowardThe (?p, ?r) → lowRisk (?p, ?a)
∧ hasAwayFromThe (?p, ?r) → lowRisk (?p, ?a)

3. Low risk: The situation involves a low level of danger. It is inferred according to the following rules:
Pedestrian (?p) ∧ Road (?r) ∧ hasHighSpeed (?p, ?s) ∧ objectOnTheRoadSide (?p, ?rs)
∧ hasAwayFromThe (?p, ?r) → lowRisk (?p, ?a)
Pedestrian (?p) ∧ Road (?r) ∧ hasMediumSpeed (?p, ?s) ∧ objectOnTheRoadSide (?p, ?rs)
∧ hasTowardThe (?p, ?r) → lowRisk (?p, ?a)
Pedestrian (?p) ∧ Road (?r) ∧ hasLowSpeed (?p, ?s) ∧ objectOnTheRoadSide (?p, ?rs)
∧ hasAwayFromThe (?p, ?r) → lowRisk (?p, ?a)

4. No risk: The situation involves no danger. It is inferred according to the following rule:
Pedestrian (?p) ∧ Road (?r) ∧ hasNoSpeed (?p, ?s) ∧ objectOnTheRoadSide (?p, ?rs) → noRisk (?p, ?a)

where p, r, re, rs, and a, represents the pedestrian, road, road edge, road side and assessment respectively.

This work was conducted using the Protégé resource[15]. Pellet reasoner [16] was used to check the consistency of the ontology, and SPARQL query was used to query in the testing stage.
3. Experimental evaluation

In this article, the evaluation of the ontology proposed in Section 2 is focused on the pedestrian safety part of the ontology; evaluation of the complete ontology will be carried out in future work. To assess our framework, we study the output of its reasoning facility when applied to real-life road scenes and compare it against ground truth. Furthermore, this output is discussed with respect to the ontology’s entities which contributed to the reasoning output. Towards this effort, we created a dataset comprising six videos featuring pedestrian behaviour in road scenes with various degrees of risk. All videos were taken from YouTube. The initial resolution of the videos varied and the frame rate was between 25-30 fps. We resized the resolution of all video frames to 640x480. All videos were captured from right-hand drive vehicles and correspond to the driver’s perspective, with legal and safety speed limits for each road type. Ground truth for the dataset, i.e. classification of each frame according to the risk concealed in the scene it illustrates to the classes no risk, low risk, medium risk and high risk was provided by two independent observers. Experiments were run on a PC with Intel i7-2600@3.40GHz CPU and 16GB of RAM running Windows 7 64-bit.

In each frame, three attributes are estimated for each pedestrian: speed, location and direction. First of all, we detect the pedestrians. There are many methods to detect pedestrians in the scene [17]. Real time video segmentation methods such as the evolving Gaussian mixture model [18] can also be used to detect pedestrians and extract semantic scene information like the road type (e.g. motorway, urban road, off-road etc.) [13]. While these methods offer good accuracy, in practice they do not guarantee perfect detection rate. Since our purpose here is to evaluate the proposed ontology, in this article pedestrians are detected manually. For this task, we developed marking software using Matlab. The inclusion of a fully automatic pedestrian detection and tracking facility in our framework will be pursued in future work. Once the pedestrians are detected, their location in the scene, speed and direction are estimated. Fig. 4 shows how these features are extracted from frames captured by a monocular camera. The distance between the centers of a pedestrian bounding box in frames $t$ and $t-1$ is estimated. This distance represents the pedestrian’s displacement between two consecutive frames and is taken as the speed of the pedestrian in terms of pixel per frame. Pedestrian speeds are classified into four classes as shown in Eq.1:

$$\text{Pedestrian Speed class} = \begin{cases} 
\text{High Speed} & \text{Speed} > \text{high} \_ \text{thresh} \\
\text{Medium Speed} & \text{low} \_ \text{thresh} > \text{Speed} \geq \text{high} \_ \text{thresh} \\
\text{Low Speed} & 0 > \text{Speed} \geq \text{low} \_ \text{thresh} \\
\text{No Speed} & \text{Speed} = 0
\end{cases} \quad (1)$$

In our study the thresholds defining low and high speed are empirically set at 3 and 6 pixels per frame, respectively. The measurements for these three attributes, which correspond to key scene entities, are fed to the ontology’s reasoning tool, which evaluates the degree of risk in the scene.

Experimental results in terms of percent classification accuracy for the six videos of our dataset are given in Fig. 5 where it is shown that the proposed ontology tool can assess the risk in the road scenes of our dataset with high accuracy. Results are reported for two hypotheses for estimating the pedestrian’s position with respect to the road. The first hypothesis takes into account the centre of the pedestrian’s bounding box and the second the vertical edge of the pedestrian’s bounding box which results in higher risk (vertical edges hypothesis). For example, if the first vertical edge is located on the road and the second on the pavement, the first edge is used. The vertical edges hypothesis offers higher classification accuracy (98.3%) than the centre of bounding box hypothesis (94.6%) for our dataset.

Representative examples of risk assessment from the dataset are presented in Fig. 6. Two of those examples are described in detail here. In Fig. 6a the object of interest is the pedestrian and the object’s location is the road. The pedestrian’s speed was estimated at 2.6 pixels per frame. According to Eq.1, this speed is classified as low. The object’s direction is 90 degrees with respect to the car driver’s perspective. The ontology tool inferred that the situation poses a high level of risk. The key feature that influences the decision is pedestrian location. In the example illustrated in Fig. 6b, the pedestrian’s speed is 5.9 pixels per frame, which, according to the Eq.1, is classified as medium speed. The object’s direction is 90 degrees with with respect to the car driver’s perspective. The ontology tool inferred that this scene does not pose risk.

In Fig. 7 we plot the output of the ontology’s inference tool over time for a video from our dataset together with the extracted features. There are four key events in this video. We explain each of these key events individually.
Fig. 5: Experimental results.

Fig. 6: Risk assessment examples.
• At $K_0$, the pedestrian ($P$) is waiting on the road side ($rs$) with no speed and direction. Therefore, according to the defined rules in section 2.3, the ontology tool inferred that the situation does not conceal risk.

• At $K_1$, the pedestrian ($P$) on the road side ($rs$) has started walking with low speed toward the road ($r$). Therefore, according to the proposed rules in section 2.3 the ontology tool inferred that the situation poses medium risk.

• At $K_2$, the pedestrian ($P$) on the road side ($rs$) is walking with high speed toward the road ($r$). Therefore, according to the proposed rules in section 2.3 the ontology tool inferred that the situation contains medium level of risk.

• At $K_3$ and $K_4$ the pedestrian ($P$) on the road ($r$) is walking with high speed. Therefore, according to the defined rules in section 2.3 the ontology tool inferred that the situation poses a high level of risk.

We notice that, for the key events $K_3$ and $K_4$, the ontology’s reasoning tool inferred the same level of risk, although the pedestrian’s speed is different in each event. This is due to an important property that appears in both events, which is on the road. According to the defined rules, when a pedestrian appears on the road, the situation poses a high level of risk, regardless of the pedestrian’s speed. For the key events $K_1$ and $K_2$, the ontology tool inferred the same level of risk as well: in this case, the common key feature between the two events is the direction of the pedestrian toward the road, regardless of the pedestrian’s speed. On the contrary, the role of speed is more important when comparing events $K_0$ and $K_1$, as it is the only factor that influences the output of the ontology tool.

4. Conclusions and future work

This article proposes a novel ontology which tackles the problem of automatic risk assessment in unpredictable road traffic environments. A framework for video-based assessment of the degree of risk in a road scene encompassing the above ontology is also presented in this paper. Unlike previous work in situational awareness, where several types of sensors were used simultaneously, the proposed framework uses as input video captured by a single monocular video camera. This yields the advantage that the required information is acquired in an efficient and inexpensive manner. Furthermore, there is no assumption that the road users obey the traffic rules when building the ontology; thus, the proposed ontology tool is designed to tackle the general, unconstrained problem of interpreting unpredictable road traffic. The evaluation of the proposed framework focuses on scenarios in which risk results from pedestrian behaviour. The framework’s performance is assessed on a dataset comprising real-world videos illustrating pedestrian...
movement. The experimental results showed that the proposed framework can accurately assess risk resulting from pedestrian behaviour in road scenes.

The future work will expand the assessment of the framework to other factors contributing to risk in road scenes, such as road type, environmental conditions and incoming traffic. Furthermore, certain components of the framework, such as the pedestrian tracker, will be automated.

5. Acknowledgment

We would like to thank Ahmed Abdulhadi Al Moadhen, PhD student at Cardiff University for his technical assistance in the implementation of the ontology.

References