

# Video-based Situation Assessment for Road Safety

---

Thesis submitted to Cardiff University in candidature for the degree  
of Doctor of Philosophy

Mahmud Abdulla Muhammad



Cardiff University  
2016

## **DECLARATION**

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed..... (candidate) Date .....

## **STATEMENT 1**

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ..... (candidate) Date .....

## **STATEMENT 2**

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit references.

Signed ..... (candidate) Date .....

## **STATEMENT 3**

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate) Date .....

## **STATEMENT 4**

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, after expiry of a bar on access approved by the Graduate Development Committee.

Signed..... (candidate) Date .....

---

---

# ABSTRACT

In recent decades, situational awareness (SA) has been a major research subject in connection with autonomous vehicles and intelligent transportation systems. Situational awareness concerns the safety of road users, including drivers, passengers, pedestrians and animals. Moreover, it holds key information regarding the nature of upcoming situations. In order to build robust automatic SA systems that sense the environment, a variety of sensors, such as global positioning systems, radars and cameras, have been used. However, due to the high cost, complex installation procedures and high computational load of automatic situational awareness systems, they are unlikely to become standard for vehicles in the near future.

In this thesis, a novel video-based framework for the automatic assessment of risk of collision in a road scene is proposed. The framework uses as input the video from a monocular video camera only, avoiding the need for additional, and frequently expensive, sensors. The framework has two main parts: a novel ontology tool for the assessment of risk of collision, and semantic feature extraction based on computer-vision methods.

The ontology tool is designed to represent the various relations between the most important risk factors, such as risk from object and road environmental risk. The semantic features related to these factors

---

are based on computer vision methods, such as pedestrian detection and tracking, road-region detection and road-type classification. The quality of these methods is important for achieving accurate results, especially with respect to video segmentation. This thesis, therefore, proposes a new criterion of high-quality video segmentation: the inclusion of temporal-region consistency. On the basis of the new criteria, an online method for the evaluation of video segmentation quality is proposed. This method is more consistent than the state-of-the-art method in terms of perceptual-segmentation quality, for both synthetic and real video datasets. Furthermore, using the Gaussian mixture model for video segmentation, one of the successful video segmentation methods in this area, new online methods for both road-type classification and road-region detection are proposed.

The proposed vision-based road-type classification method achieves higher classification accuracy than the state-of-the-art method, for each road type individually. Consequently, it achieves higher overall classification accuracy. Likewise, the proposed vision-based road-region detection method achieves high performance accuracy compared to the state-of-the-art methods, according to two measures: pixel-wise percentage accuracy and area under the receiver operating characteristic (*ROC*) curve (*AUC*).

Finally, the evaluation performance of the automatic risk-assessment framework is measured. At this stage, the framework includes only the assessment of pedestrian risk in the road scene. Using the semantic information obtained via computer-vision methods, the framework's performance is assessed for two datasets: first, a new dataset proposed in Chapter 7, which comprises six videos, and second, a dataset com-

---

prising five examples selected from an established, publicly available dataset. Both datasets consist of real-world videos illustrating pedestrian movement. The experimental results show that the proposed framework achieves high accuracy in the assessment of risk resulting from pedestrian behaviour in road scenes.

---

---

# ACKNOWLEDGEMENTS

First, I am grateful to Allah, who made it possible for me to complete my journey towards earning my PhD. Second, I would like to thank my supervisors, Dr. Yulia Hicks and Professor Rossitza Setchi, for their continuous support during my PhD study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in every phase of the research and in writing this thesis. I could not have hoped for a better advisors and mentors.

In addition to my advisors, I would like to thank Dr. Ioannis Kaloskampus for his continuous support and encouragement during my PhD study. He constantly provided good ideas and excellent guidance.

I would also like to thank the Kurdistan Regional Government - Human Capacity Development Program (KRG-HCDP), my supervisors Dr. Yulia Hicks and Professor Rossitza Setchi, and Cardiff University (Engineering School and Student support center) for their financial support during this study.

Last but not the least, my sincere thanks also goes to my wife Zainab Rasul Aziz for her endless love, support and prayers. Also, I thank my beloved children, Zhir, Yad and Shayan. Finally, I would like to convey my sincere gratitude and appreciation to my family; my mother, father, grandfather, brothers and sisters have supported me spiritually throughout my writing of this thesis and all my life in general.

---

---

# LIST OF ACRONYMS

<b>3D</b>	Three-dimensional
<b>AS</b>	Combining appearance and structure method [Sturgess et al., 2009]
<b>AUC</b>	Area under the curve
<b>ABox</b>	Assertional box
<b>ANN</b>	Artificial neural network
<b>BU</b>	Bottom-up, data-driven road-detection method [Alvarez et al., 2014]
<b>CamVid</b>	Cambridge-driving labelled video database
<b>CNN</b>	Convolutional neural network
<b>CRF</b>	Conditional random field
<b>CMU</b>	Carnegie Mellon University database
<b>CPU</b>	Central processing unit
<b>DSIFT</b>	Dense scale invariant feature transform
<b>F</b>	F-measure
<b>GPS</b>	Global positioning system

---

<b>GMMs</b>	Gaussian mixture models
<b>GCE</b>	Global consistency error
<b>GCI</b>	Global consistency index
<b>GISs</b>	Geographical information systems
<b>Hog</b>	Histograms of oriented gradients
<b>ITS</b>	Intelligent transportation system
<b>ITI</b>	Intelligent transportation infrastructure
<b>K-NN</b>	K-nearest neighbour
<b>KITTI</b>	Karlsruhe Institute of Technology and Toyota Technological Institute
<b>LIDAR</b>	Light detection and ranging
<b>LBP</b>	Local binary patterns
<b>MRF</b>	Markov random field
<b>OWL DL</b>	Web ontology language with description logics
<b>OWL</b>	Web ontology language
<b>PC</b>	Personal computer
<b>PRI</b>	Probabilistic Rand index
<b>P</b>	Precision
<b>PR</b>	Proposed method
<b>RAM</b>	Random access memory

---

<b>RDF</b>	Resource description framework
<b>R</b>	Recall
<b>ROC</b>	Receiver operating characteristic
<b>RS</b>	Road scene segmentation method [Alvarez et al., 2012]
<b>RWA</b>	Random walker algorithm
<b>SA</b>	Situational awareness
<b>SVM</b>	Support vector machine
<b>SOM</b>	Self-organizing map
<b>SUN</b>	Scene understanding database
<b>SWRL</b>	Semantic web rule language
<b>SPARQL</b>	Simple protocol and resource description framework (RDF) query language
<b>STR</b>	State of the art
<b>TBox</b>	Terminological box
<b>TD</b>	Top-down, data-driven road-detection method [Alvarez et al., 2014]
<b>VI</b>	Variation of information
<b>XML</b>	Extensible Markup Language

---

---

# List of Figures

2.1	Assessing the road situation.	16
2.2	CamVid dataset, examples of manually segmented (ground truth) and machine segmentation.	23
2.3	An example of intraregion uniformity.	24
2.4	Subregions used in [Tang and Breckon, 2011, Mioulet et al., 2013].	29
2.5	Difficult cases for methods in [Tang and Breckon, 2011, Mioulet et al., 2013].	30
2.6	Road structure types.	35
3.1	Ontology structure for automatic risk assessment.	41
3.2	Object attributes portion of the ontology structure.	42
3.3	Pedestrian speed and direction calculation.	48
3.4	Hypotheses for estimating pedestrian position.	50
3.5	Experimental risk-assessment accuracy.	51
3.6	Risk-assessment examples.	52
3.7	Event-based graphical illustration of the ontology results.	54

---

4.1	Example of unsupervised boundary detection.	64
4.2	Examples of various defects in synthetic video.	69
4.3	Sample of the synthetic video dataset.	70
4.4	Visual comparison of the six aspects of segmentation quality and ground truth.	70
4.5	Result of: (a) $F$ and $F'$ ; (b) $P$ and $P'$ .	71
4.6	Recall $R$ accuracy results on real video.	72
4.7	Evaluation results of intraregion uniformity ( $U$ ) and temporal-region consistency ( $C$ ) on real video.	73
4.8	Segmentation evaluation results using precision $P$ and updated precision $P'$ .	74
4.9	Segmentation evaluation results using $F$ and $F'$ .	75
4.10	Final overall results of real video evaluation using $F$ and $F'$ .	76
5.1	Process of building road-type models.	81
5.2	Pipeline of the classification process.	83
5.3	Road-type classification results using both methods.	88
5.4	Road-class labels.	89
5.5	Visual comparison of motorway example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.	90
5.6	Visual comparison of off-road example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.	90

---

5.7	Visual comparison of trunk example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.	91
5.8	Visual comparison of urban example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.	91
5.9	Examples of misclassified frames by the proposed method, and correctly classied frames by STR [Mioulet et al., 2013]	91
6.1	The threefold road model.	97
6.2	Road-detection pipeline.	97
6.3	Process of building the road and nonroad models with the EvoGMM algorithm.	98
6.4	Example of probability map and global prior mask.	101
6.5	Process of generating the initial road region.	104
6.6	Example of the initial road region and its comparison with the ground truth.	104
6.7	Example of road region produced from boundary refinement using superpixels.	105
6.8	Example of road region produced using boundary refinement and the prior road shape model.	108
6.9	Example on the effect of the thresholds and their comparison with the ground truth.	109
6.10	Example of estimated road region using region-growing and its comparison with the ground truth.	110

---

6.11	Evaluation of road-detection methods and comparison between the state-of-the-art methods and the proposed method.	113
6.12	Qualitative evaluation of the proposed method for the CamVid dataset: (a) Original; (b) Initial road region; (c) Superpixel boundary refinement; (d) Prior road shape model; (e) Region growing; (f) Superimposition of the output of proposed method on the original frame.	113
7.1	CamVid dataset and two examples of road detection using the proposed method.	117
7.2	CamVid dataset and two examples of pedestrian detection using the Viola-Jones algorithm [Viola and Jones, 2001a].	119
7.3	Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for case 1 footage 1.	122
7.4	Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for case 1 footage 2.	123
7.5	Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for (a) case 1 footage 3 and (b) case 1 footage 4.	125
7.6	Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for (a) case 1 footage 5 and (b) case 1 footage 6.	126

---

7.7	Proposed dataset for pedestrian behaviour in road scenes; accuracy results for (a) case 1 footage 1, (b) case 1 footage 2, (c) case 1 footage 3, (d) case 1 footage 4, (e) case 1 footage 5 and (f) case 1 footage 6.	127
7.8	CamVid dataset; results of comparison with ground truth for (a) case 2 footage 1 and (b) case 2 footage 2.	129
7.9	CamVid dataset; results of comparison with ground truth for (a) case 2 footage 3 and (b) case 2 footage 4.	130
7.10	CamVid dataset; results of comparison with ground truth for case 2 footage 5.	131
7.11	Proposed dataset for pedestrian behaviour in road scenes; accuracy results for (a) case 2 footage 1, (b) case 2 footage 2, (c) case 2 footage 3, (d) case 2 footage 4 and (e) case 2 footage 5.	132
7.12	Accuracy results for (a) all video footages from the proposed dataset, (b) all video footages from CamVid dataset, (c) both case studies and (d) overall risk assessment for each case study.	134
A.1	SPARQL query, test rule number 1.	163
A.2	SPARQL query, test rule number 2.	163
A.3	SPARQL query, test rule number 3.	164
A.4	SPARQL query, test rule number 4.	164
A.5	SPARQL query, test rule number 5.	164
A.6	SPARQL query, test rule number 6.	165

---

A.7 SPARQL query, test rule number 7.	165
A.8 SPARQL query, test rule number 8.	165
A.9 SPARQL query, test rule number 9.	166
A.10 SPARQL query, test rule number 10.	166
A.11 SPARQL query, test rule number 11.	167
A.12 SPARQL query, test rule number 12.	167
A.13 SPARQL query, test rule number 13.	168
A.14 SPARQL query, test rule number 14.	168
A.15 SPARQL query, test rule number 15.	169
A.16 SPARQL query, test rule number 16.	169
A.17 SPARQL query, test rule number 17.	170
A.18 SPARQL query, test rule number 18.	170
A.19 SPARQL query, test rule number 19.	171
A.20 SPARQL query, test rule number 20.	171
A.21 SPARQL query, test rule number 21.	172
A.22 SPARQL query, test rule number 22.	172
A.23 SPARQL query, test rule number 23.	173
B.1 MATLAB query, main menu to choose the entities.	175
B.2 MATLAB query indicating that the Vulnerable and Ob- ject on the road buttons are selected.	176
B.3 MATLAB query, test rule number 1.	176
B.4 MATLAB query, test rule number 2.	177
B.5 MATLAB query, test rule number 3.	177

---

B.6	MATLAB query, test rule number 4.	177
B.7	MATLAB query, test rule number 5.	178
B.8	MATLAB query, test rule number 6.	178
B.9	MATLAB query, test rule number 7.	179
B.10	MATLAB query, test rule number 8.	179
B.11	MATLAB query, test rule number 9.	180
B.12	MATLAB query, test rule number 10.	180
B.13	MATLAB query, test rule number 11.	181
B.14	MATLAB query, test rule number 12.	181
B.15	MATLAB query, test rule number 13.	182
B.16	MATLAB query, test rule number 14.	182
B.17	MATLAB query, test rule number 15.	183
B.18	MATLAB query, test rule number 16.	183
B.19	MATLAB query, test rule number 17.	184
B.20	MATLAB query, test rule number 18.	184
B.21	MATLAB query, test rule number 19.	185
B.22	MATLAB query, test rule number 20.	185
B.23	MATLAB query, test rule number 21.	186
B.24	MATLAB query, test rule number 22.	186
B.25	MATLAB query, test rule number 23.	187

---

---

# List of Tables

1.1	Summary of the video sequences from the [Chen and Corso, 2010] dataset	7
1.2	Dataset summary for road type classification	8
1.3	Summary of CamVid dataset for road detection	8
1.4	Summary of proposed dataset for pedestrian behaviour in road scenes	9
1.5	Summary of CamVid dataset for pedestrian behaviour in road scenes	10
2.1	Categorisation of datasets based on scene types	33
3.1	Summary of the proposed dataset for pedestrian behaviour in road scenes	51
4.1	Summarise the applied criteria, measures and metrics	59
5.1	Dataset summary for road type classification	86
5.2	Classification results in terms of classification accuracy percentage	89
5.3	Confusion matrix of the proposed method	89

---

5.4	Confusion matrix of [Mioulet et al., 2013]	89
6.1	Summary of CamVid dataset	111
7.1	Summary of proposed dataset for pedestrian behaviour in road scenes	124
7.2	Summary of CamVid dataset for pedestrian behaviour in road scenes	133

---

---

# CONTENTS

<b>ABSTRACT</b>	<b>iii</b>
<b>ACKNOWLEDGEMENTS</b>	<b>vi</b>
<b>LIST OF ACRONYMS</b>	<b>vii</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xvii</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Motivation	1
1.2 Challenges	2
1.3 Aims and objectives	5
1.4 Datasets	6
1.5 Contributions	10
1.6 Publications from this study	12
1.7 Submitted article	12
1.8 Thesis overview	13
	<b>xix</b>

---

<b>2</b>	<b>LITERATURE REVIEW</b>	<b>14</b>
2.1	Situational awareness in road scenes	14
2.2	Video segmentation and evaluation	18
2.2.1	Video segmentation methods	19
2.2.2	Evaluation methods	22
2.3	Scene understanding	26
2.3.1	Road-type classification	27
2.3.2	Road detection	29
2.4	Summary	35
<b>3</b>	<b>ONTOLOGY-BASED FRAMEWORK FOR RISK AS-</b>	
	<b>SESSMENT IN ROAD SCENES USING VIDEOS</b>	<b>37</b>
3.1	Risk-assessment method	39
3.1.1	Ontologies	39
3.1.2	Structure of the proposed ontology	40
3.1.3	Rule-based cases	43
3.2	Experimental evaluation	46
3.3	Summary	54
<b>4</b>	<b>A NEW METHOD FOR EVALUATION OF VIDEO</b>	
	<b>SEGMENTATION QUALITY</b>	<b>56</b>
4.1	Proposed criteria and metrics	57
4.1.1	Intraregion uniformity and homogeneity	59
4.1.2	Temporal-region consistency	60

---

4.1.3	Boundary stability and accuracy	60
4.2	Proposed evaluation method	61
4.2.1	Intraregion uniformity	61
4.2.2	Region consistency	65
4.2.3	Boundary assessment	66
4.2.4	Combining metrics	66
4.3	Evaluation and results	68
4.3.1	Synthetic data	68
4.3.2	Real video data	68
4.3.3	Results on synthetic video	70
4.3.4	Results on real video	72
4.4	Summary	77
<b>5</b>	<b>EVOLVING GMMS FOR ROAD-TYPE CLASSIFICATION</b>	<b>78</b>
5.1	Online video segmentation	79
5.2	Building the road-type model	81
5.3	Classification	83
5.4	Experimental results	86
5.5	Summary	92
<b>6</b>	<b>ONLINE ROAD DETECTION</b>	<b>94</b>
6.1	Method overview	95
6.2	Building the threefold road model	97

---

6.2.1	The road and nonroad model	98
6.2.2	The prior road shape model	99
6.2.3	The naive outlier filtering mask	100
6.3	Road detection	101
6.3.1	Initial road region generation	101
6.3.2	Road-boundary refinement with superpixels	103
6.3.3	Handling light effects with the prior road shape model	105
6.3.4	Unusual road shape handling with the region growing method	108
6.4	Experiments and results	110
6.5	Summary	113
<b>7</b>	<b>VIDEO-BASED ASSESSMENT OF THE DEGREE OF RISK IN A ROAD SCENE</b>	<b>115</b>
7.1	Semantic feature extraction based on computer vision methods	116
7.1.1	Road region detection	117
7.1.2	Pedestrian detection and tracking	118
7.2	Speed, location and direction calculation	119
7.3	Data combination	119
7.4	Assessment of pedestrian risk	120
7.5	Risk assessment results and evaluation	120
7.5.1	Case study 1: proposed dataset	121

---

7.5.2	Case study 2: CamVid dataset	128
7.5.3	Overall results	133
7.6	Summary	135
<b>8</b>	<b>CONCLUSIONS AND FUTURE WORK</b>	<b>136</b>
8.1	Conclusions	136
8.2	Limitations	139
8.3	Future work	140
	<b>REFERENCES</b>	<b>141</b>
	<b>APPENDIX A</b>	<b>162</b>
	<b>APPENDIX B</b>	<b>174</b>

## INTRODUCTION

### 1.1 Motivation

The last two decades have witnessed considerable improvement in the field of transportation infrastructure. Recent developments in autonomous vehicles have resulted in intelligent automobiles and an integrated transport system, which uses a wide range of technologies, namely, communication, control, vehicle sensing and electronics [Singh and Gupta, 2015, Wang et al., 2015a]. This is due to the integration of these technologies into a manufacturing model, which provides intelligent services for specific aspects of transport and traffic management [Yan et al., 2012], such as traffic flow, congestion, optimum routes, safety, complexity and cost [Singh and Gupta, 2015].

However, in spite of these improvements in the field of intelligent automobiles, major traffic problems have continued to increase. This is essentially due to the considerable increase in the number of vehicles [Singh and Gupta, 2015]. The traffic problems include human safety, traffic flow and congestion. In the design of intelligent transportation systems (ITSs), the problems related to human safety are paramount. Therefore, researchers have recently shown an increased interest in the issue of safety as it relates to intelligent automobiles [Wang et al., 2006,

Spehr et al., 2011].

The safety of intelligent automobiles is a comprehensive notion, related mainly to the safety of road users, that is, drivers, passengers, pedestrians and animals. In designing intelligent vehicles, many safety equipments have been used to protect drivers and passengers when accidents happen; these offer passive protection for drivers and passengers [Wang et al., 2006]. Moreover, active hazard perception is the key to understanding the nature of any upcoming situation and preventing accidents [Wang et al., 2006]. Knowing the level of risk of collision and situational awareness (SA) will help to prevent accidents, which can be assessed on the basis of the perception information.

Of course, risk assessment to avoid collision with vulnerable road users and situational awareness (SA) will improve the safety of intelligent systems in terms of safety. SA can be defined briefly as ‘knowing what’s going on’ [Endsley, 1995] or ‘keeping track of what is going on around you in a complex, dynamic environment’ [Vincenzi et al., 2004]. SA is considered to be a key process in autonomous driving.

## 1.2 Challenges

Achieving SA fully is crucial in complex, dynamic scene environments. Notably, recent advances in autonomous vehicle technology raise the important problem of automatic SA in road scenes. In order to build robust automatic SA systems that can sense the environment, a variety of sensors, such as global positioning systems (GPSs), radars and cameras have been used.

In recent years, there has been an increasing interest in using several types of sensors simultaneously. Although the presence of multiple

sensors offers rich information [Bengler et al., 2014], due to high costs, complex installation procedures and high computational load, the use of multiple sensors will not become standard for vehicles in the near future. Certain sensors, such as ultrasonic, radar and laser, may also suffer from interference [Yang and Zheng, 2015]. Therefore, a good option is to use a monocular camera, as it is an efficient sensor in terms of cost and richness of information [Liu et al., 2013]. The main purpose of using these sensors is to provide important information related to the entities in the scene; by processing this information, the sensors can identify semantic features of the scene entities, which is an important step in the achievement of scene understanding and object recognition.

The recognition of important scene entities around a vehicle is crucial to assessing the risk of collision in a given road scene [Ess et al., 2008]. However, object recognition does not provide sufficient information to evaluate the situation in terms of safety, because the behaviour of these objects is also important. For instance, while driving, seeing a child on the road is a riskier situation than seeing a child on the pavement. Here, in the former instance, the behaviour of the child has more meaning with respect to safety. In the past, to solve the situation assessment problem, a variety of techniques were used (e.g. Bayesian networks and ontology). Recently, ontologies have been used successfully to efficiently model complex interactions between entities in road scene environments and to represent a wide variety of behaviours.

Because video data is high-dimensional data, feature extraction is an effective way to deal with it. The purpose of feature extraction is to reduce the amount of data by measuring certain ‘features’, ‘attributes’ or ‘properties’, and then passing them to the next step of processing

[Duda et al., 2012]. In many application areas, semantic features, which express the existence or nonexistence of semantic entities in the scene, are more pragmatic than other kinds of features.

Extracting semantic features is a crucial issue in computer vision and video processing. Video segmentation is an effective process for reading and interpreting basic digital video data with respect to its semantic content [Ngan and Li, 2011], which means subdividing images into nonoverlapping, meaningful segments [Dey et al., 2010, Morris et al., 1986]. There are many different approaches and algorithms for video segmentation; hence, their evaluation is also important for assessing the quality of segmentation results. Nonetheless, little research has focused on the evaluation of video segmentation quality.

Semantic segmentation aims to reduce the semantic gap between the low-level features and high-level semantics. Achieving high-level semantics is an essential component in the analysis and understanding of the content of the scene. The utilisation of the semantic information of both the scene entities and their behaviours results in a reasonable and fair assessment of the situation.

As outlined above, in spite of tremendous efforts and excellent progress in the areas of SA, video sensors, intelligent automobiles and transport systems, the ‘automatic safety’ problem persists [Bengler et al., 2014].

The hypothesis of this thesis is as follows: *If the semantic features of all key entities in the road scene can be obtained from video frames by using computer vision methods, they can be organised into an ontology structure that encodes their hierarchy, relations and interactions. Using video-based features as sources of information about the key entities, the*

*ontology tool can then infer the behaviour of the entities and the degree of risk of collision in a given scene. This service can be standardised for all types of vehicles.*

### **1.3 Aims and objectives**

In this thesis, a framework for video-based assessment of the risk of collision in a road scene is considered. The framework is built around a novel ontology that encompasses the key entities in the road scene and encodes their hierarchy, relations and interactions. The framework uses as input the video from a monocular video camera only, avoiding the need for additional, and frequently expensive, sensors. Hence, some video-processing techniques are needed to support the framework's understanding and analysis of the video data.

Over the last two decades, many approaches have been proposed to solve the challenges related to video segmentation, but many problems persist. The challenges associated with segmentation are not unique problems in the field, but the challenges faced by the evaluation of segmentation are an important focus of research. Evaluation of segmentation quality enables researchers to select the best video segmentation method for use in real-world applications. The current methods of video segmentation evaluation consider the boundaries of the segmentations without taking into account region interiors and consistency throughout the video. Thus, a robust method for evaluation of segmentation quality is required to assess which algorithm provides more accurate segmentation.

In this study, online video segmentation, as an early processing step in video analysis, is combined with other computer-vision methods to

interpret the semantic content of video data. This process effectively reduces the semantic gap between the low-level features and the high-level semantics. The semantic information obtained from computer-vision methods represent key entities in the road scene. The measurements of the key scene entities are then fed to the ontology's reasoning tool, which evaluates the degree of risk of collision in the scene.

The aim of this research is to build a novel video-based framework of SA for road safety. The specific objectives are as follows:

1. Development of a video-based framework for risk assessment in road scenes that takes into account all factors that influence the risk assessment of the scene
2. Development of an algorithm for the evaluation of video segmentation quality, and using it to choose a suitable online video-segmentation algorithm
3. Development of new methods to identify key entities in the road scene, such as road types and road regions
4. Assessment of the video-based framework for the degree of risk in a road scene

## 1.4 Datasets

In this thesis, several datasets are used to evaluate the proposed methods and to conduct the experiments. In the following paragraphs, the details of each dataset are discussed:

1. For the evaluation of video segmentation quality, two types of datasets are used, synthetic dataset and real dataset. Two syn-

thetic videos with the same length are created. The first video depicts several semantic objects (a circle, for example) moving from place to place. The second video represents different defects in the segmentation of the first video. Moreover, the ‘correctly segmented’ frames between the ‘defective segmentations’ are inserted to represent inconsistent temporal segmentations. The real videos are selected from a publically available dataset [Chen and Corso, 2010], and the lengths of the selected video sequences are listed in Table 1.1.

Table 1.1: Summary of the video sequences from the [Chen and Corso, 2010] dataset

Sequences	Num. frames
Bus	85
Container	86
Garden	81
Ice	80
Soccer	69
Stefan	76
All frames	919

2. For evaluation of the road type classification, a dataset is built as follows: The videos used for the urban road model are taken from [Brostow et al., 2008], whereas the videos used for the rest of the road types are taken from YouTube. The lengths of the selected video sequences are listed in Table 1.2.

Table 1.2: Dataset summary for road type classification

Sequences	Num. frames
Off-road	1000
Motorway	1000
Urban road	1000
Trunk road	1000

3. For evaluation of the proposed road detection method, the publicly available CamVid dataset [Brostow et al., 2008] is used. The dataset includes daytime and dusk sequences captured from right-hand drive vehicles and correspond to the driver’s perspective. The resolution of the frames is 960X720 pixels. The lengths of the selected video examples are listed in Table 1.3.

Table 1.3: Summary of CamVid dataset for road detection

Sequences	Num. frames
EX_1_0001TP	124
EX_2_0006R	101
EX_3_0016E_1	127
EX_4_0016E_2	178
Ex_5_05VD	171
All frames	701

4. The performance of the risk assessment framework is assessed on

two datasets: first, a proposed new dataset which comprises six videos, all of which are taken from YouTube, second, five video examples selected from the CamVid dataset [Brostow et al., 2008]. Both datasets comprise real-world videos illustrating pedestrian movement. The lengths of video examples of both case studies are listed in Tables 1.4 and 1.5.

Table 1.4: Summary of proposed dataset for pedestrian behaviour in road scenes

Sequences	Num. frames
Case 1 footage 1	83
Case 1 footage 2	75
Case 1 footage 3	59
Case 1 footage 4	105
Case 1 footage 5	64
Case 1 footage 6	131
All frames	517

Table 1.5: Summary of CamVid dataset for pedestrian behaviour in road scenes

Sequences	Num. frames
Case 2 footage 1	49
Case 2 footage 2	433
Case 2 footage 3	49
Case 2 footage 4	299
Case 2 footage 5	89
All frames	919

## 1.5 Contributions

The findings from this study make several contributions to the current state of art:

1. A novel ontology structure for risk assessment in road scenes using videos is proposed, which tackles the problem of automatic risk assessment in unpredictable road traffic environments. The structure includes the factors that influence the risk assessment of the scene. Furthermore, the ontology structure does not assume that road users obey the traffic rules.
2. New criteria of high-quality video-segmentation are proposed, and a new evaluation method based on these criteria is designed, which can be used both for supervised and unsupervised evaluation. A synthetic video test set is created specifically for the purpose of evaluating the performance of the proposed method.

3. A new online vision-based road-type classification method is proposed. The method uses video captured by a single video camera and takes into account the visual information of the whole scene by segmenting the video frames into temporally consistent frame segments.
4. A new online model-based road-detection method is proposed. The method uses video captured by a single video camera and is followed by two steps of region refinement. The advantages of many different classifiers are combined to boost the confidence levels of the road-region pixels.
5. A video-based framework for assessing the degree of risk of collision in a road scene is proposed. The framework consists of several steps. The first, semantic-feature extraction, is based on computer-vision methods. The second step is the calculation of speed, location and direction. The third step is data combination, and the fourth is assessment of pedestrian risk.

## 1.6 Publications from this study

The following publications are based on the work presented in this thesis:

1. M. A. Mohammad, I. Kaloskampis, Y. Hicks, “New Method for Evaluation of Video Segmentation Quality,” *10<sup>th</sup> International Conference on Computer Vision Theory and Applications (VIS-APP)*. Berlin, Germany. March 2015.
2. M. A. Mohammad, I. Kaloskampis, Y. Hicks, “Evolving GMMs for road-type classification,” *IEEE International Conference on Industrial Technology (ICIT)*. Seville, Spain. March 2015.
3. M. A. Mohammad, I. Kaloskampis, Y. Hicks, R. Setchi “Ontology-based framework for risk assessment in road scenes using videos,” *19<sup>th</sup> International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, (KES)*. Singapore. 7–9 Sept. 2015.

## 1.7 Submitted article

M. A. Mohammad, I. Kaloskampis, Y. Hicks, “Video-based road detection using evolving GMMs, shape priors and region enhancement,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), the 7th International Workshop on Computer Vision in Vehicle Technology (CVVT)*. Las Vegas, Nevada. US. 26 June–1 July 2016.

## 1.8 Thesis overview

The remainder of this thesis is structured as follows:

- *Chapter 2*: reviews the existing research related to the work presented in this thesis.
- *Chapter 3*: proposes ontology-based risk assessment in road scenes using videos.
- *Chapter 4*: presents a new method of evaluation of video segmentation quality.
- *Chapter 5*: describes a new online vision-based road-type classification method.
- *Chapter 6*: describes a new online model-based road-detection method.
- *Chapter 7*: describes the video-based evaluation of the ontology framework.
- *Chapter 8*: presents the conclusions and limitations of the thesis, with suggestions for future work.

# LITERATURE REVIEW

In this chapter, the existing research related to the work presented in this thesis is reviewed, and the relevant techniques and methods that have an impact on problem solving are described. This chapter is structured as follows: SA in road scenes is discussed in Section 2.1. The general background of and related work on video segmentation and evaluation methods are investigated in Section 2.2. This investigation identifies and analyses the video segmentation methods that can be used in this framework. Section 2.2 also explains the challenges facing current and future research on video segmentation and evaluation methods. Scene understanding is then reviewed in Section 2.3; scene understanding is a central research topic in computer vision and has been used in many real-vision applications. The findings are summarised in Section 2.4.

### **2.1 Situational awareness in road scenes**

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, the automobiles are capable of providing many intelligent services with respect to the various forms of transport and

traffic management [Yan et al., 2012], such as traffic flow, congestion, optimum routes, safety, complexity and coast [Singh and Gupta, 2015].

Safety for drivers and other road users requires an important intelligent service in the design of autonomous vehicles, which can be achieved by investigating the SA in order to determine the degree of risk of collision in the scene, given a number of sensor measurements. The notion of SA can be defined briefly as ‘knowing what’s going on,’ ‘the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future’ [Endsley, 1995] or ‘keeping track of what is going on around you in a complex, dynamic environment’ [Vincenzi et al., 2004]. SA is considered a key process in autonomous driving.

Recognition of important scene objects around the vehicle is crucial when assessing the risk of collision in a given road scene [Ess et al., 2008]. However, object recognition does not provide sufficient information to evaluate the situation with respect to safety, because the behaviour of these objects is also important. Figure 2.1 shows two scenes featuring the same objects. In Figure 2.1a, the pedestrian is on the road and the situation is, therefore, riskier than in Figure 2.1b, where the pedestrian is on the pavement and moving away from the road.

At the same time, certain environmental factors influence the risk assessment of the scene, such as visibility conditions (fog, haze pollution and light), weather, traffic signs, road type and road quality [Pollard et al., 2013]. Therefore, the assessment of risk of collision 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 Figure 2.1, for instance, the interactions between the pedestrian, the road and the pavement influence the degree of risk of collision.

Researchers have employed several different methods to solve this problem. Platho et al. [Platho et al., 2012] 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, because there are no direct interactions between different sets, this method may have problems propagating the effect of events from one set to another. Schamm and Zöllner [Schamm and Zollner, 2011] used a knowledge-based framework that takes into account interactions between entities to solve the problem. Vacek et al. [Vacek et al., 2007] addressed 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 [Platho et al., 2012].

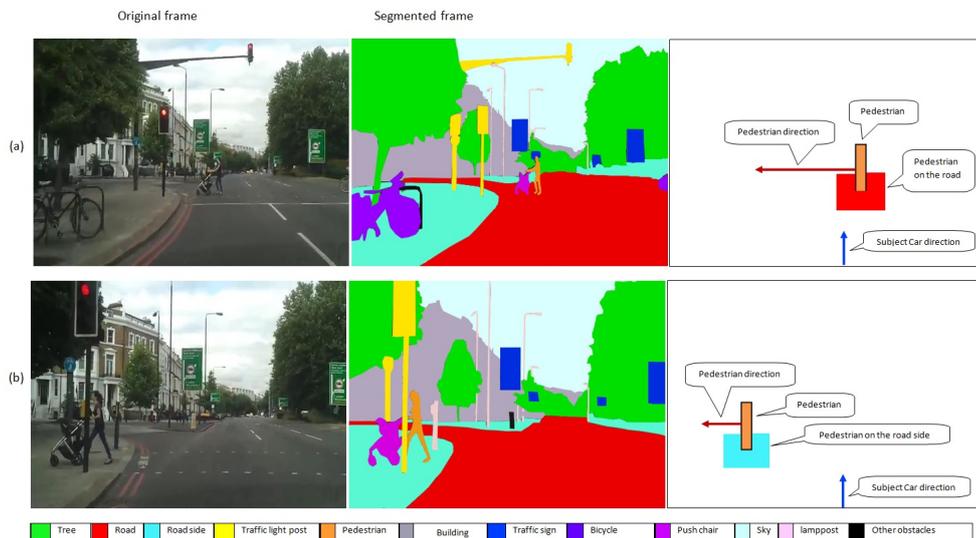


Figure 2.1: Assessing the road situation.

However, there are several methods for semi-automatic ontology learning [Maedche, 2012], such as pattern-based [Hearst, 1992] or definition-

based [Navigli and Velardi, 2010] methods. Manually built ontologies have been used successfully to efficiently model complex interactions between entities in road-scene environments and to represent a wide variety of behaviours without stability issues. Hülsen et al. [Hülsen et al., 2011] proposed an ontology-based situation-description method for traffic intersections. Pollard et al. [Pollard et al., 2013] presented an ontology for situation assessment for automated ground vehicles that takes into account 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. Armand et al. [Armand et al., 2014] proposed an ontology-based SA framework that 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 environments.

It is also worth noting that in the frameworks described above, 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 costs and complexity reasons, the use of multiple sensors will not become standard for vehicles in the near future. Certain sensors, such as ultrasonic, radar and laser, may also suffer from interference problems [Yang and Zheng, 2015].

## 2.2 Video segmentation and evaluation

Regardless of the challenges associated with the other sensors, single monocular video cameras are becoming standard for vehicles because of their relatively low cost and the richness of information they provide. Thus, investigating the SA provided by this type of sensor is vital. Because video data is high-dimensional data, many effective techniques have been developed to reduce its dimensionality by measuring certain ‘features,’ ‘attitudes,’ and ‘properties’. Researchers have categorised these features into two types. In [Nixon and Aguado, 2008], the image features were divided into two categories:

- Low-level features, which can be defined as local properties or pixel-based attitudes that are extracted from an image, irrespective of shape information or spatial attitude, or as image-level descriptors that characterise the image content
- High-level features, which can be defined as global properties or region-level descriptors of an image’s shape or spatial attitude.

Both types of features have been used extensively to segment images/frames by passing them to the next step of processing [Duda et al., 2012] and segmenting them semantically. The aim of segmentation is to find nonoverlapping semantic regions of an image/frame.

Segmentation is a crucial issue in computer vision and image/video processing, where it is defined as the subdividing of images into meaningful segments [Dey et al., 2010, Morris et al., 1986]. This technique has become an effective process for reading and interpreting the semantic content of basic digital images/video data [Ngan and Li, 2011]. Therefore, it plays a central role in image/video analysis and under-

standing, and supports applications like object recognition, image coding and image indexing [Allili et al., 2010, Goldberger and Greenspan, 2006a, Huang et al., 2009]. Because many approaches to and algorithms for image/video segmentation have been developed, it is important to evaluate the quality of their segmentation results. Nonetheless, little research has focused specifically on the evaluation of video segmentation quality.

### 2.2.1 Video segmentation methods

Video segmentation has been an important research area for decades, and researchers have proposed many different approaches to providing high-quality segmentation. These approaches have been categorised into different groups. The feature-based category includes approaches based on appearance [Vazquez-Reina et al., 2010, Brendel and Todorovic, 2009, Grundmann et al., 2010, Lezama et al., 2011, Kaloskampis and Hicks, 2014, Charron and Hicks, 2010], motion [Galasso et al., 2011, Brox and Malik, 2010, Shi and Malik, 2000] and combinations of feature cues [Galasso et al., 2012, Levinshtein et al., 2010, Cheng and Ahuja, 2012, DeMenthon and Megret, 2002, Greenspan et al., 2002, Kannan et al., 2005, Kumar et al., 2008, Paris, 2008, Lee et al., 2011, Ochs and Brox, 2011].

Although previous studies have reported that combinations of feature cues can provide better cues and lead to better results [Galasso et al., 2012], this approach involves a tradeoff between scene content types and feature types. Clearly, motion features can provide better clues for dynamic scenes than static scenes, whereas appearance features are not significantly affected by changes between two consecu-

tive frames as affected by perspective, illumination and contrast in the scene. It is, therefore, possible to control this tradeoff on the basis of detected changes between two consecutive frames, as demonstrated by the technique of updating GMM parameters proposed by [Kaloskampis and Hicks, 2014].

Moreover, video segmentation approaches can be categorised on the basis of the techniques used, such as graph-based models [Grundmann et al., 2010], mean shift [DeMenthon and Megret, 2002, Greenspan et al., 2002, Kannan et al., 2005, Kumar et al., 2008, Paris, 2008], the Gaussian mixture model [Kaloskampis and Hicks, 2014, Charron and Hicks, 2010] layered models [Kannan et al., 2005, Kumar et al., 2008] and spectral clustering [Galasso et al., 2012, Arbelaez et al., 2009, Brox and Malik, 2010, Shi and Malik, 2000]. Again, the approaches based on these techniques can be classified according to user interaction phenomena: supervised [Vogel et al., 2006] or unsupervised [Kaloskampis and Hicks, 2014] or time-dimension phenomena: online [Kaloskampis and Hicks, 2014] or real-time and offline [Grundmann et al., 2010].

Because it is difficult to obtain general public segmentation, and because segmentation methods are considered application-oriented methods [Ravi et al., 2016, Ngan and Li, 2011], choosing a preferred method for a specific application requires better categorisation. However, categorisation based on a single aspect provides valuable information regarding the approaches proposed in the literature, whereas, categorisation based on multiple aspects provides better clues and may provide better support for the selection of an appropriate approach for a specific application.

Multi-aspect categorisation classifies video segmentation methods

on an application-specific basis. Each application has unique limitations and goals [Ngan and Li, 2011]; therefore, applications do not require the same quality or type of segmentation [Lobato Correia and Pereira, 2004]. Correia and Pereira [Lobato Correia and Pereira, 2004] grouped the applications of video segmentation into four broad categories:

- real-time (online) non-user interactive scenario; applications that belong to this scenario are identified by the real-time process without user interaction. This scenario includes the applications related to direct broadcasting, video surveillance, and online video coding [Tang and Breckon, 2011]
- real-time (online) user interactive scenario; applications that belong to this scenario are identified by the real-time process along with extra support, like user interaction. This scenario includes the applications related to video-conferences [Askar et al., 2004]
- offline nonuser interactive scenario; this scenario is related to the applications that require automatic algorithms, without considering the real-time implementation, such as video indexing and offline video coding for automatic segmentation [Izquierdo and Ghanbari, 2002]
- offline user interactive scenario; this scenario is related to applications that require supervised algorithms without considering the real-time implementation, including applications related to offline video coding, such as video summarization [Chang, 2003]

The first scenario has been identified as a fully automatic segmentation solution [Lobato Correia and Pereira, 2004], which is required

in many vision applications. Moreover, the proposed framework is a real-world application, and it is supposed to work online without user interaction. Therefore, the framework proposed in this thesis seeks to choose a video segmentation method for the first scenario; to this end, the online and unsupervised method proposed in [Kaloskampis and Hicks, 2014] is chosen.

### 2.2.2 Evaluation methods

Segmentation is an important stage in image/video analysis and understanding. Because many different approaches to and algorithms for image/video segmentation have been developed, it is important to evaluate the quality of their segmentation results. Nonetheless, little research has focused specifically on the evaluation of video segmentation quality. Researchers have divided the evaluation methods into three classes [Zhang et al., 2008, Correia and Pereira, 2003].

**Subjective evaluation** is the evaluation process in which human observers quantify the quality of segmentation results on the basis of visual description. This is a complicated and time-consuming process, and the results vary from one observer to another.

**Supervised evaluation** is the evaluation process in which a segmented image/frame (Figure 2.2 c), is compared to a manually segmented (ground truth) reference image/frame (Figure 2.2 b). Producing ground truth images is also a time-consuming process, and it involves a certain degree of disagreement between different people.

**Unsupervised evaluation**, also known as stand-alone evaluation or empirical goodness evaluation, works automatically without any extra requirements such as ground truth images. The methods in this

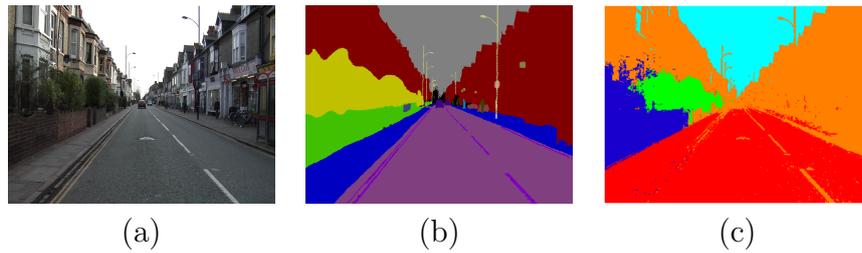


Figure 2.2: CamVid dataset, examples of different segmentation methods: (a) Original frame; (b) Manually segmented (ground truth); (c) segmented example using [Kaloskampis and Hicks, 2014].

evaluation class use only low-level features and do not incorporate semantic information. The most important characteristic of unsupervised methods is that they can be used to control the parameters of online video segmentation in real-time applications [Zhang et al., 2008].

Most of the evaluation methods are subjective or related to specific applications. The majority of the proposed objective evaluation methods fall into the category of supervised evaluation, while the area of unsupervised evaluation has received the least attention [Zhang et al., 2008]. Evaluation is usually based on several criteria, each of which considers the quality of the segmentation from a different perspective. A number of researchers have considered which aspects of segmentation quality should be evaluated. In the remainder of this section, the existing criteria and metrics will be reviewed.

Levine and Nazif [Levine and Nazif, 1985] suggested that to design a measure for evaluating the quality of image segmentation, it is necessary to consider the following: (1) uniformity within regions, (2) contrast across regions and (3) provision for lines and texture. Figure 2.3 shows two examples of high-quality segmentation.

Haralick and Shapiro [Haralick and Shapiro, 1985] proposed four criteria for the evaluation of image segmentation: (1) regions must be

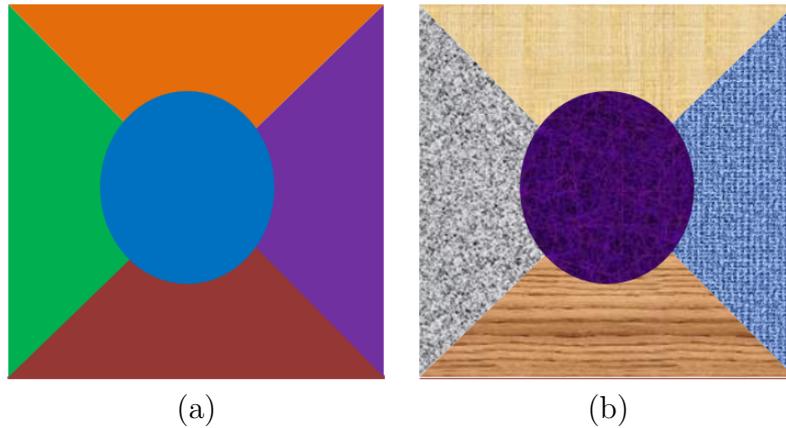


Figure 2.3: An example on intraregion uniformity: (a) Uniform regions; (b) Regions with uniform textures.

uniform and homogeneous, (2) adjacent regions should have significant differences with respect to the characteristic on which they are uniform, (3) region interiors should be simple and without holes (Figure 2.3) and (4) boundaries should be smooth and accurate. Most of the previously developed evaluation methods and metrics incorporate the above criteria, either directly or indirectly [Levine and Nazif, 1985, Liu and Yang, 1994, Borsotti et al., 1998, Chen and Wang, 2004, Zhang et al., 2004, Chabrier et al., 2006].

[Zhang et al., 2008] classified the evaluation methods according to the criteria proposed in [Haralick and Shapiro, 1985]. The classification also covers unsupervised metrics proposed for the evaluation of image and video segmentation. They concluded that these criteria had become the de facto standard for unsupervised evaluation of image segmentation. They concluded that the first two criteria were more characteristic than semantic and hence incorporated the first and second criteria into their work. Zhang et al. [Zhang et al., 2008] conducted a comparative evaluation of different approaches and concluded that previously developed unsupervised approaches for the evaluation

of image-segmentation methods are insufficient for the comparison of segmentation produced by different algorithms.

The criteria discussed above have been applied to the evaluation of the quality of image segmentation. The previously developed unsupervised methods for the evaluation of video segmentation methods [Correia and Pereira, 2003, Erdem et al., 2004] are limited and not designed for general-purpose applications: the former method involves the manual labeling of data, and the latter is designed for evaluating video object segmentation and tracking algorithms. Likewise, the metric proposed in [Gelasca and Ebrahimi, 2006] is based on spatial and temporal accuracy and designed for evaluating video object segmentation.

In addition to the methods described above, there are several popular supervised evaluation methods based on image/frame boundaries as opposed to regions. The boundary precision-recall metric is used in [Martin et al., 2001] as a supervised metric for the evaluation of image segmentation. Galasso et al. [Galasso et al., 2013] introduced the volume precision-recall metric for evaluation of video segmentation quality. Xu et al. [Xu and Corso, 2012] proposed 3D volumetric quality metrics to evaluate super-voxel methods, which they based on boundaries without taking into account region uniformity and consistency.

The current state of the art in the evaluation of video segmentation quality can be summarised as follows: (1) there are no established criteria for evaluation of overall video segmentation as opposed to image segmentation or video object segmentation, (2) there are a limited number of unsupervised evaluation methods of video segmentation, and they are not designed for overall video segmentation and (3) supervised evaluation methods of video segmentation consider the boundaries of

the segmentations without taking into account region interiors.

Therefore, the evaluation of video segmentation quality requires new criteria. Based on the new criteria, an online method for the evaluation of video segmentation quality can be built, which takes into account the characteristics of both boundaries and regions.

### 2.3 Scene understanding

The comprehensive understanding of the video content of the scene plays a crucial role, and it can be exploited to sense the environment. This can be achieved by understanding the video frames by labelling the frame regions. Clearly, an essential component of this understanding is inferring semantic and high-level information from the scene [Liu et al., 2014], which has a fundamental impact on the performance of many intelligent vehicle applications [Spehr et al., 2011] and of many computer-vision applications, such as browsing, retrieval, object recognition [Gökalp and Aksoy, 2007] and scene classification [Choi et al., 2014].

Vision-based intelligent vehicle applications cover a variety of smart services provided by autonomous vehicles; these applications are capable of inferring semantic information from the road scene. Each application can be considered a step towards the understanding of the road scene: lane detection, traffic sign recognition, obstacle detection, pedestrian detection and tracking, road detection and road type classification.

Road detection is an important application for both robotic and autonomous vehicles [Wang et al., 2015b], and it helps researchers to understand a given situation in terms of safety and the crucial aspect

of safe access to the road for road users. In addition, each road type requires a specific behaviour from road users. Thus, road detection and road-type classification are both important applications, and both are included in the framework proposed in this thesis.

### 2.3.1 Road-type classification

Scene classification is an important and challenging topic in the field of scene understanding [Choi et al., 2014]. Many research contributions have been made in both indoor and outdoor scene classification [Li and Guo, 2014] based on local features, such as the histogram of textons [Leung and Malik, 2001], the bag-of-words (BoW) [Csurka et al., 2004, Sivic and Zisserman, 2003], hypergraph-based modelling [Choi et al., 2014], a combination of local and global information such as bag-of-regions [Gökalp and Aksoy, 2007] and adaptive active learning [Li and Guo, 2014].

In addition, road-type classification, as a specific type of scene classification, is an important step towards road-scene understanding. Road scene understanding is required in a variety of applications in the areas of SA and fully automated or semiautomated driving [Tang and Breckon, 2011]. In such applications, exploiting domain knowledge information is the key. However, extracting domain knowledge information from the perception of the road environment is a major challenge in autonomous systems [Miranda Neto et al., 2013], and it requires high-quality image/video processing methods [Mioulet et al., 2013].

Over the last three decades, many research contributions have been made in visual navigation [Bonin-Font et al., 2008]; nonetheless, building robust methods remains an important problem [Miranda Neto et al.,

2013]. In recent years, a considerable amount of research has focused on the use of different types of sensors, but in terms of cost and richness of information, using a monocular camera is preferable [Liu et al., 2013]. Examples of work in this area include road-environment classification [Mioulet et al., 2013, Tang and Breckon, 2011], road detection [Alvarez and Lopez, 2011, Broggi and Berte, 1995], road marking [Kheyrollahi and Breckon, 2012], road-sign detection and recognition [Piccioli et al., 1996], on-road sign analysis [Eichner and Breckon, 2008], off-road environment classification [Jansen et al., 2005], and highway lane detection [Melo et al., 2006].

The work presented in [Tang and Breckon, 2011, Mioulet et al., 2013] focuses on the problem of road-type classification. There are three main steps in each of these approaches: region selection, feature extraction and preparation, and classification. Both methods select three subregions of interest from the frames of the road video sequences—road, road edge and roadside—but use different features and classifiers in the second and third steps. The method in [Tang and Breckon, 2011] extracts colour, texture and edge-derived features and applies k-nearest neighbor (k-NN) and artificial neural network (ANN) classification approaches, whereas the method in [Mioulet et al., 2013] extracts Gabor texture features and uses the random forests classifier [Breiman, 2001]. The method in [Mioulet et al., 2013] achieved higher accuracy classification than the method in [Tang and Breckon, 2011].

In both methods [Tang and Breckon, 2011, Mioulet et al., 2013], as illustrated in Figure 2.4, three subregions were selected as the interest regions for the driving environment: road, road edge and roadside. The properties of these three regions are captured and used as key

information during classification. However, there is no guarantee that the subregions will capture all key information.

In addition, there are specific cases in which the subregions are unlikely to contain the key information, such as when the car turns left or right, or is driven on a rough road. Figure 2.5 provides some examples of the difficulties associated with both methods [Tang and Breckon, 2011, Mioulet et al., 2013].

To overcome such problems, it is necessary to take into account all regions in the frame. One way to achieve this is to use an online video segmentation method and then compare the detected segments to those usually found in certain types of roads.

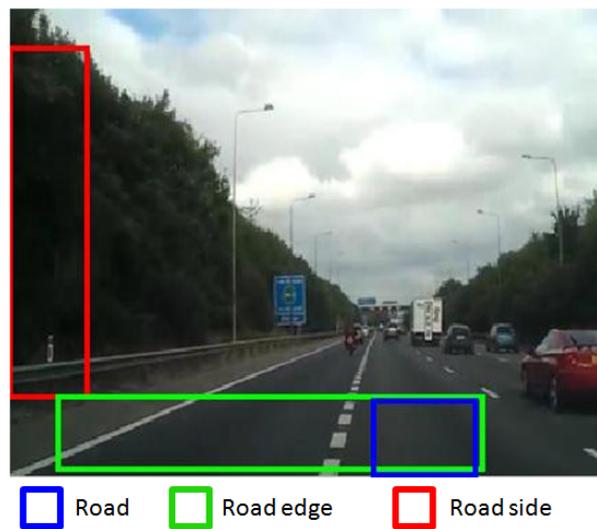


Figure 2.4: Subregions used in [Tang and Breckon, 2011, Mioulet et al., 2013].

### 2.3.2 Road detection

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



Figure 2.5: Difficult cases for methods in [Tang and Breckon, 2011, Mioulet et al., 2013].

by these sensors, they are capable of determining the travel route and identifying important scene objects, such as traffic signs and obstacles. An important problem in the design of autonomous vehicles is road detection, as it provides an important information cue to sense the environment and eases applications such as path finding and planning, object tracking, anomaly detection and situation assessment.

Vision-based approaches typically use colour as the main low-level feature for road detection, as texture is dependent on reliable shape patterns parallel to the road direction and increases computational costs [Álvarez et al., 2014]. However, important environmental challenges such as colour variation, shadows and lighting conditions pose problems to colour-based road detectors [Wang et al., 2015b], hence additional information is required to improve the detection accuracy.

For roads that are designed in accordance with design guidelines and standards, road structure can be used as a cue to improve the system's performance [Han et al., 2012, Jiang et al., 2014]. The drawback of such approaches is that they cannot operate reliably in unstructured road scenes (Figure 2.6).

Typical road geometries can also be exploited to enhance the performance of road detection. For instance, in [Alvarez et al., 2009] geometries like left turn, straight and T-like junction are learned offline and a scene classifier selects the most probable profile for a given input frame. Such approaches lose accuracy in certain circumstances, e.g. in cluttered scenes [Álvarez et al., 2014].

Other methods employ a combination of techniques to enhance the performance of road detection. For example, appearance-based and motion features were utilised in [Sturgess et al., 2009]. As this method

relies on specific training samples, it has difficulties coping with images that are significantly different to the training data. Colour plane fusion (CPF) and convolutional neural networks (CNNs) were combined in [Alvarez et al., 2012]. This work assumes that the bottom part of the video frame captures the road-region, which, however, cannot be always guaranteed in practice [Alvarez et al., 2014].

Prior knowledge regarding the road-shape has also been used to improve the performance of road detection, e.g. [He et al., 2013]. The road-shape is typically learnt from training data or past frames and is used to impose restrictions regarding the detected road area in the input frame. Such methods may face problems when the road-region in the input frame is significantly different than the models in the learnt road-shape database.

The shortcomings of vision-based approaches have led researchers to include additional sensors in their systems alongside traditional cameras, such as stereo cameras [Guo and Mita, 2009, Guo et al., 2012, Bertozzi and Broggi, 1998, Wang and Fremont, 2013], thermal cameras [Pelaez et al., 2015], radar [Hu et al., 2014, Feng et al., 2012, Ma et al., 2000], LIDAR [Prochazka, 2014], GPS and GIS [Álvarez et al., 2014] and multi-sensor solutions [Han et al., 2012]. Although the additional sensors improve the road detection accuracy, due to high cost, complex installation procedures and high computational load they are currently not close to becoming standard for vehicles. Moreover, certain sensors, such as ultrasonic, radar, laser and GPS, may additionally suffer from interference problems.

Accordingly, the existence of a visual sensor is gradually becoming standard for modern vehicles, with an increasing number of vehicles

being equipped with dashboard cameras. Therefore, an online vision-based road detection method is crucial and a method to handle problems in the detected road-region caused by shadows, illuminations and unusual road-shapes.

In addition, for the evaluation of the proposed methods in this area, there are two standard and publically available datasets, the CamVid dataset [Brostow et al., 2008] and the KITTI dataset [Fritsch et al., 2013]. There are some other publically available datasets, such as the Alvarez dataset [Alvarez and Lopez, 2011], the SUN dataset [Xiao et al., 2010], the CMU dataset [Cmu, 1997], the SIFT Flow dataset [Liu et al., 2011a] and the Stanford background dataset [Gould et al., 2009]. As explained in Table 2.1, these datasets can be categorised by their scene types, such as urban area [Brostow et al., 2008, Fritsch et al., 2013], trunk road [Alvarez and Lopez, 2011], driving range [Xiao et al., 2010], trunk road and off-road with different shadow and illumination conditions [Cmu, 1997] and a variety of outdoor scenes [Liu et al., 2011a, Gould et al., 2009].

Table 2.1: Categorisation of datasets based on scene types

Scene type	Datasets
Urban area	[Brostow et al., 2008, Fritsch et al., 2013]
Trunk road	[Alvarez and Lopez, 2011]
Driving range	[Xiao et al., 2010]
Trunk road and off-road	[Cmu, 1997]
Variety of outdoor scenes	[Liu et al., 2011a, Gould et al., 2009]

Obviously, having different datasets offers more information on a variety of road-scene types (motorway, urban road, trunk road and off-road) under different conditions (weather and light), and provides a variety of views and aspects than an individual dataset. Therefore, robust methods should be evaluated using datasets with a wide-range of possibilities.

Although tremendous efforts have been made in this area and a range of methods have been proposed for detecting road areas, the majority of them have been evaluated on their own datasets as opposed to standard datasets. Therefore, except for the methods that have been evaluated with standard datasets, it is difficult to rank the methods in terms of their effectiveness. The CamVid dataset [Brostow et al., 2008], which is widely used as a standard dataset by the state-of-the-art methods, was the preferred dataset for this study.

Moreover, the implementation codes of most of these methods are not publically available or are limited to a specific company (e.g. [Yao et al., 2015]). Sometimes, the performance of some of the methods decreases considerably when applied on another dataset. Therefore, more robust research approaches should be taken, and the approaches should be evaluated qualitatively using an existing and publically available dataset. At the same time, it is important to help other researchers by providing program codes and approach guidelines.



Figure 2.6: Road structure types; (a) structured road type; (b) unstructured road type

## 2.4 Summary

This chapter has reviewed the previous research that is relevant to this thesis, as well as the relevant techniques that have been used to address the problems considered in this thesis. The findings are as follows:

1. An ontology structure for risk assessment in road scenes using videos is needed. It should achieve the following goals:
  - address the problem of automatic risk assessment in unpredictable road traffic environments
  - include the factors that influence the risk assessment of the scene
  - include no assumption that road users obey the traffic rules
  - incorporate a risk-assessment framework for validation with real video data
2. For evaluation of video segmentation quality, the following are needed:
  - a new criterion of high-quality video segmentation: temporal-region consistency

- 
- a new evaluation method based on the new criterion
3. Road type classification is a step toward road-scene understanding, and the road types have their own individual impact in assessing risk of collision. Therefore, considering the visual information of the whole scene, it is important to improve the accuracy of the road type classifications.
  4. Road detection is of essential importance for both robotic and autonomous vehicles; any improvement in this area is valuable. Improvement can be achieved by exploiting the advantages of different models and classifiers.
  5. There is no dataset designed for evaluating video-based risk-assessment methods; a new dataset in this area is needed.

# ONTOLOGY-BASED FRAMEWORK FOR RISK ASSESSMENT IN ROAD SCENES USING VIDEOS

Recent advances in autonomous vehicle technology pose the important problem of automatic risk assessment in road scenes. As explained in Chapter 2, determining the degree of risk of collision in a given road scene is an important aspect of the design of autonomous vehicles. Many types of sensors have been used to provide valuable semantic information about road scenes. Although this technology is important, it represents only part of the equation. As explained in Figure 2.1, from the safety point of view, the challenges associated with the role and behaviour of objects in road scenes extend beyond their recognition.

This chapter addresses the problem of automatic risk assessment by proposing a novel ontology tool for the assessment of risk of collision in unpredictable road traffic environments, because the tool does not assume that road users always obey the traffic rules. A framework for video-based assessment of risk of collision in a road scene encompassing

the ontology tool is also presented in this chapter. The framework uses as input the video from a monocular video camera only, avoiding the need for additional, and frequently expensive, sensors. The key entities in the road scene (vehicles, pedestrians, environment objects, etc.) are organised manually into an ontology that encodes their hierarchy, relations and interactions. The ontology tool infers the degree of risk of collision in a given scene using as knowledge video-based features related to the key entities. In this specific application, due to the lack of training data and the performance level of the methods, it is preferable to manually build the ontology.

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 of collision in the scene. The experimental results indicate that the proposed framework is capable of accurately assessing risk resulting from pedestrian behaviour in various road scenes.

The contributions of this chapter are as follows:

1. a novel ontology tool for the assessment of risk of collision in unpredictable road traffic environments
2. a framework for video-based assessment of risk of collision in a road scene encompassing the proposed ontology
3. a dataset consisting of real-world videos illustrating pedestrian movement

This chapter is organised as follows. In Section 3.1, the proposed risk-assessment method is discussed. Experimental results are given in Section 3.2. Finally, the main conclusions of the chapter are summarised

in Section 3.3.

### 3.1 Risk-assessment method

As explained in the literature review (Section 2.1), assessing the degree of risk of collision 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 chapter, a novel ontology-based framework for assessing the degree of risk of collision in a road scene is proposed. This ontology is designed to address risk of collision related to several factors, such as risk from objects (vehicles, pedestrians, cyclists, etc.), environmental risk (weather conditions and visibility conditions) and road environmental risk (road quality, road traffic signs and road types). The proposed method will be explained in more depth below.

#### 3.1.1 Ontologies

In philosophy, ontology is defined as an ‘account of existence’ [Gruber, 1993]. In computer engineering, the definition of ontology is the ‘specification of a conceptualisation’ [Gruber, 1993]. More specifically, ontology is a hierarchical definition of the terms and the relationships between them, resulting in a formal representation of knowledge that is understandable by humans and computers [Armand et al., 2014]. An ontology-based framework consists of a terminological box (TBox), which includes concepts, role definitions and axioms, and an assertional box (ABox), which includes instances of concepts and the roles of such instances [Pollard et al., 2013, Armand et al., 2014].

### 3.1.2 Structure of the proposed ontology

As explained in Section 2.1, risk assessment of the road scene is influenced by certain environmental factors, such as visibility conditions (fog, haze pollution and light), weather, traffic signs, road type and road quality. Thus, automatic risk assessment involves the processing of a plethora of information arising from several factors and their interaction. To manage these factors and their interactions, a novel ontology-based framework for assessing the degree of risk of collision in a road scene is proposed. The framework is shown in Figure 3.1. It consists of three main classes that correspond to factors contributing to risk: collision risk, environmental risk and road environmental risk. In the next paragraph, each of these classes is discussed individually.

The risk factor classes comprise 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, the structure of the risk factor classes is discussed:

1. Collision risk: The role of this class is to provide detailed information regarding the object attributes, so that the degree of risk of collision can be assessed from the type and behaviour 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 subclasses of object direction) and object type. Moreover, the object-type class consists of two subclasses: Vulnerable (with three subclasses representing different types of vulnerable) and Vehicle. Finally, object location (with three subclasses of different locations).

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. The weather-conditions class contains two subclasses: normal weather condition and bad weather condition (with three subclasses representing different types of bad weather conditions). Visibility conditions consist of two subclasses: normal visibility and reduced visibility (with five subclasses representing different types of reduced visibility conditions).
3. Road environmental risk: The role of this class is to provide rich information about the road environment based on the interact with the other factors in the proposed ontology, the risk level can be assessed. This class consists of two subclasses: road qual-

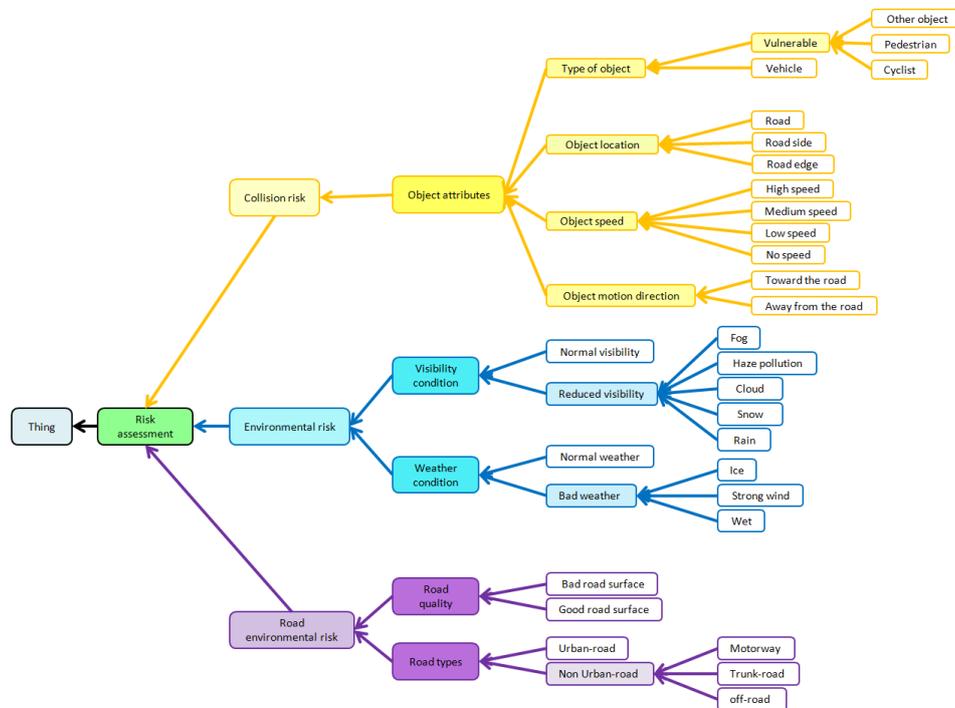


Figure 3.1: Proposed ontology structure for automatic risk assessment in road scenes.

ity (with a subclass of bad road surface and good road surface) and road type (with two subclasses of urban road and non-urban road). The non-urban road consists of three subclasses: motorway, trunk road and off-road.

An object property is the binary relation between two classes. Here, 15 object properties are defined based on the necessity of the relations, namely *highRisk*, *mediumRisk*, *lowRisk*, *noRisk*, *hasHighSpeed*, *hasMediumSpeed*, *hasLowSpeed*, *hasNoSpeed*, *hasAwayFrom*, *hasTowardThe*, *objectOnTheRoad*, *objectOnTheRoadEdge*, *objectOnTheRoadSide*, *badRoadSurface*, and *goodRoadSurface*.

In this structure, to assess the risk level of the *RiskAssessment*, only one of the properties among *highRisk*, *mediumRisk*, *lowRisk*, and *noRisk* must be inferred. Again, only one of the speed properties among *hasHighSpeed*, *hasMediumSpeed*, *hasLowSpeed* and *hasNoSpeed* must be inferred, and these properties specify the speed type, *ObjectSpeed*, of the *vulnerable*. The object-motion direction property, *ObjectMotionDirection*, of the *vulnerable*, according to the *Road* observer, is inferred by *hasAwayFrom* and *hasTowardThe*. Finally, the intersection of the object location, *ObjectLocation* with the object *vulnerable* is inferred on the basis of one of the properties among *objectOnTheRoad*, *objec-*

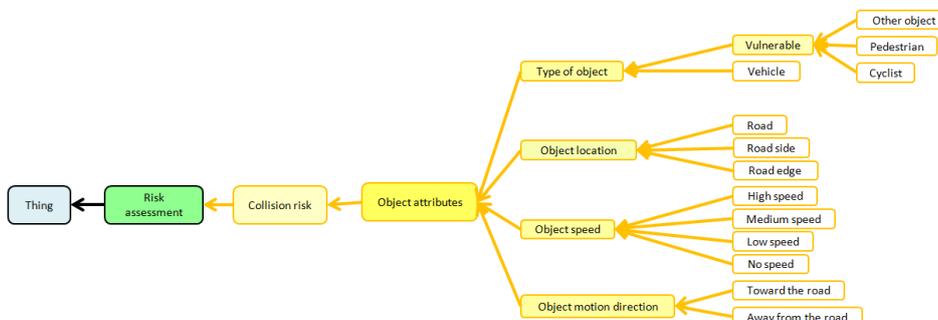


Figure 3.2: Object attributes portion of the ontology structure.

*tOnTheRoadEdge* and *objectOnTheRoadSide*.

### 3.1.3 Rule-based cases

In this section, 23 rule-based cases for the proposed structure are defined. These rules are based on human knowledge and the information from the risk factor classes. The rules are formed in the semantic web rule language (SWRL) [Horrocks et al., 2004]. This format is a generic language from the Web Ontology Language (OWL), and it is based on a combination of the OWL DL and OWL Lite sublanguages [Horrocks et al., 2004].

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

$$\text{ReducedVisibility}(?y) \rightarrow \text{isHighRisk}(?y, ?a) \quad (3.1)$$

$$\begin{aligned} \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{nonUrbanRoadType}(?r, ?n) \\ \rightarrow \text{isHighRisk}(?O, ?a) \end{aligned} \quad (3.2)$$

$$\begin{aligned} \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{objectOnTheRoad}(?O, ?r) \\ \rightarrow \text{isHighRisk}(?O, ?a) \end{aligned} \quad (3.3)$$

$$\begin{aligned} \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{NormalWeather}(?c) \\ \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{badRoadSurface}(?r, ?b) \\ \wedge \text{hasAwayFromThe}(?O, ?r) \wedge \text{objectOnTheRoadEdge}(?O, ?r) \\ \rightarrow \text{isHighRisk}(?O, ?a) \end{aligned} \quad (3.4)$$

$$\begin{aligned} \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{NormalWeather}(?c) \\ \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\ \wedge \text{hasTowardThe}(?O, ?r) \wedge \text{objectOnTheRoadEdge}(?O, ?r) \\ \wedge \text{hasHighSpeed}(?O, ?s) \rightarrow \text{isHighRisk}(?O, ?a) \end{aligned} \quad (3.5)$$

$$\begin{aligned} \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{NormalWeather}(?c) \\ \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\ \wedge \text{hasTowardThe}(?O, ?r) \wedge \text{objectOnTheRoadEdge}(?O, ?r) \\ \wedge \text{hasMediumSpeed}(?O, ?s) \rightarrow \text{isHighRisk}(?p, ?a) \end{aligned} \quad (3.6)$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\wedge hasTowardThe(?O, ?r) \wedge objectOnTheRoadEdge(?O, ?r) \\
&\quad \wedge hasLowSpeed(?O, ?s) \rightarrow isHighRisk(?O, ?a) \quad (3.7)
\end{aligned}$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge BadWeather(?c) \\
&\wedge urbanRoadType(?r, ?u) \wedge objectOnTheRoadEdge(?O, ?r) \\
&\quad \rightarrow isHighRisk(?O, ?a) \quad (3.8)
\end{aligned}$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge BadWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge badRoadSurface(?r, ?b) \\
&\wedge hasTowardThe(?O, ?r) \wedge objectOnTheRoadEdge(?O, ?r) \\
&\quad \wedge hasNoSpeed(?O, ?s) \rightarrow isHighRisk(?O, ?a) \quad (3.9)
\end{aligned}$$

2. Medium risk: The situation involves a medium level of danger.

It is inferred according to the following rules:

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\wedge hasAwayFromThe(?O, ?r) \wedge objectOnTheRoadEdge(?O, ?r) \\
&\quad \wedge hasHighSpeed(?O, ?s) \rightarrow isMediumRisk(?O, ?a) \quad (3.10)
\end{aligned}$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\quad \wedge objectOnTheRoadEdge(?O, ?r) \wedge hasNoSpeed(?O, ?s) \\
&\quad \rightarrow isMediumRisk(?O, ?a) \quad (3.11)
\end{aligned}$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\wedge hasAwayFromThe(?O, ?r) \wedge objectOnTheRoadEdge(?O, ?r) \\
&\quad \wedge hasMediumSpeed(?O, ?s) \rightarrow isMediumRisk(?O, ?a) \\
&\hspace{15em} (3.12)
\end{aligned}$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\wedge hasAwayFromThe(?O, ?r) \wedge objectOnTheRoadEdge(?O, ?r) \\
&\quad \wedge hasLowSpeed(?O, ?s) \rightarrow isMediumRisk(?O, ?a) \quad (3.13)
\end{aligned}$$

$$\begin{aligned}
& \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{BadWeather}(?c) \\
& \quad \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\
& \wedge \text{hasAwayFromThe}(?O, ?r) \wedge \text{objectOnTheRoadSide}(?O, ?r) \\
& \quad \wedge \text{hasHighSpeed}(?O, ?s) \rightarrow \text{isMediumRisk}(?O, ?a) \quad (3.14)
\end{aligned}$$

$$\begin{aligned}
& \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{BadWeather}(?c) \\
& \quad \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\
& \wedge \text{hasAwayFromThe}(?O, ?r) \wedge \text{objectOnTheRoadSide}(?O, ?r) \\
& \quad \wedge \text{hasMediumSpeed}(?O, ?s) \rightarrow \text{isMediumRisk}(?O, ?a) \\
& \hspace{15em} (3.15)
\end{aligned}$$

$$\begin{aligned}
& \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{BadWeather}(?c) \\
& \quad \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\
& \wedge \text{hasAwayFromThe}(?O, ?r) \wedge \text{objectOnTheRoadSide}(?O, ?r) \\
& \quad \wedge \text{hasLowSpeed}(?O, ?s) \rightarrow \text{isMediumRisk}(?O, ?a) \quad (3.16)
\end{aligned}$$

$$\begin{aligned}
& \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{NormalWeather}(?c) \\
& \quad \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\
& \wedge \text{hasAwayFromThe}(?O, ?r) \wedge \text{objectOnTheRoadEdge}(?O, ?r) \\
& \quad \rightarrow \text{isMediumRisk}(?O, ?a) \quad (3.17)
\end{aligned}$$

$$\begin{aligned}
& \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{NormalWeather}(?c) \\
& \quad \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\
& \quad \wedge \text{objectOnTheRoadEdge}(?O, ?r) \wedge \text{hasNoSpeed}(?O, ?s) \\
& \quad \rightarrow \text{isMediumRisk}(?O, ?a) \quad (3.18)
\end{aligned}$$

3. Low risk: The situation involves a low level of danger. It is inferred according to the following rules:

$$\begin{aligned}
& \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{BadWeather}(?c) \\
& \quad \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\
& \quad \wedge \text{objectOnTheRoadSide}(?O, ?r) \wedge \text{hasNoSpeed}(?O, ?s) \\
& \quad \rightarrow \text{isLowRisk}(?O, ?a) \quad (3.19)
\end{aligned}$$

$$\begin{aligned}
& \text{Vulnerable}(?O) \wedge \text{Road}(?r) \wedge \text{NormalWeather}(?c) \\
& \quad \wedge \text{urbanRoadType}(?r, ?u) \wedge \text{goodRoadSurface}(?r, ?g) \\
& \wedge \text{hasAwayFromThe}(?O, ?r) \wedge \text{objectOnTheRoadSide}(?O, ?r) \\
& \quad \wedge \text{hasHighSpeed}(?O, ?s) \rightarrow \text{isLowRisk}(?O, ?a) \quad (3.20)
\end{aligned}$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\wedge hasAwayFromThe(?O, ?r) \wedge objectOnTheRoadSide(?O, ?r) \\
&\quad \wedge hasMediumSpeed(?O, ?s) \rightarrow isLowRisk(?O, ?a) \quad (3.21)
\end{aligned}$$

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\wedge hasAwayFromThe(?O, ?r) \wedge objectOnTheRoadSide(?O, ?r) \\
&\quad \wedge hasLowSpeed(?O, ?s) \rightarrow isLowRisk(?O, ?a) \quad (3.22)
\end{aligned}$$

4. No risk: The situation involves no danger. It is inferred according to the following rule:

$$\begin{aligned}
&Vulnerable(?O) \wedge Road(?r) \wedge NormalWeather(?c) \\
&\quad \wedge urbanRoadType(?r, ?u) \wedge goodRoadSurface(?r, ?g) \\
&\quad \wedge objectOnTheRoadSide(?O, ?r) \wedge hasNoSpeed(?O, ?s) \\
&\quad \quad \rightarrow isNoRisk(?O, ?a) \quad (3.23)
\end{aligned}$$

where  $O$ ,  $r$ ,  $re$ ,  $rs$  and  $a$  represent the Vulnerable, road, road edge, roadside and assessment, respectively.

This study was conducted using the *Protégé* resource [pro, 2015]. The Pellet reasoner [Dentler et al., 2011] was used to check the consistency of the ontology, and the SPARQL query was used for querying in the testing stage. In addition, these rules were encoded in the MATLAB function, and the results of both the SPARQL query and MATLAB query are presented in Appendix A and B, respectively.

## 3.2 Experimental evaluation

In this section, the pedestrian-safety portion of the ontology is evaluated; an evaluation of the complete ontology will be carried out in future work. To assess the proposed framework, the output of its reasoning facility when, applied to real-life road scenes, is investigated and then compared against ground truth. Furthermore, this output is discussed

with respect to the ontology’s entities that contribute to the reasoning output. Towards this effort, a dataset comprising six videos was created. This dataset consists of 517 frames of videos featuring pedestrian behaviour in road scenes with various degrees of risk. All videos, which were taken from YouTube, had the following features: good road surface, normal visibility, urban road type and normal weather condition. The initial resolution of the videos varied, and the frame rate was between 25 and 30 fps. The resolution of all video frames was resized to 640 x 480. 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. the classification of each frame according to the risk concealed in the scene, according 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, the pedestrians have to be detected. Thus, there are many methods for detecting pedestrians in a scene [Taiana et al., 2013]. Although these methods offer good accuracy, in practice they do not guarantee a perfect detection rate. Because the purpose here is to evaluate the proposed ontology, the manual detection of pedestrians is examined in this chapter. For this task, the marking software was developed using MATLAB. The inclusion of a fully automatic pedestrian detection and tracking facility in this framework will be discussed in Chapter 7. Once the pedestrians are detected, their location in the scene, speed and direction are estimated. Figure 3.3 shows

how these features are extracted from frames captured by a monocular camera. The distance between the centres 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. 3.24:

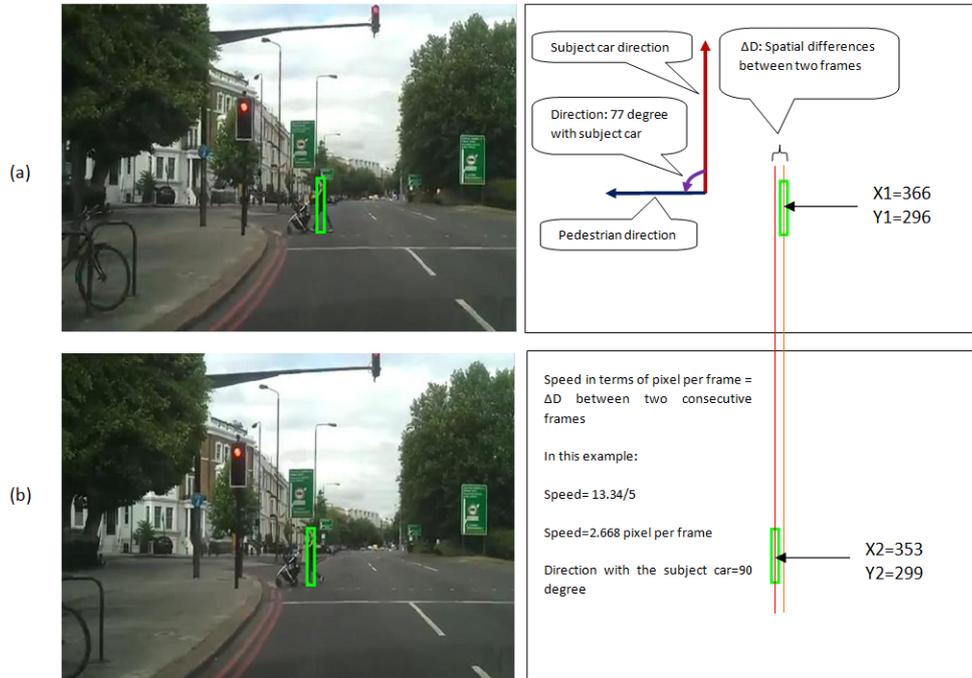


Figure 3.3: Pedestrian speed and direction calculation.

$$S^{class} = \begin{cases} H_S & S_{ped} > H_{thr} \\ M_S & L_{thr} > S_{ped} \geq H_{thr} \\ L_S & 0 > S_{ped} \geq L_{thr} \\ N_S & S_{ped} = 0 \end{cases} \quad (3.24)$$

where  $S^{class}$  is the classified speed of a pedestrian that has  $S_{ped}$  and  $H_S$ ,  $M_S$ ,  $L_S$  and  $N_S$  are the speed types high speed, medium speed, low speed and no speed, respectively. In this study, the thresholds defining

low  $L_{thr}$  and high  $H_{thr}$  speed are empirically set at 3 and 6 pixels per frame, respectively.

In this work, speed calculation is solely based on the pedestrian's displacement between two consecutive frames; the speed type is then classified using the defined thresholds. These thresholds ( $L_{thr}$  and  $H_{thr}$ ) are affected by some factors, such as the distance between the pedestrian and the camera and the relative motions between them. For example, the closer the pedestrian is to the camera, the larger is the distance between two consecutive frames; this leads to the production of a higher speed, which in turn increases the risk of collision. However this speed calculation is not strong enough to produce the speed accurately, especially in complex environments. There was no further investigation due to the time limitations of this research.

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 of collision in the scene.

Experimental results in terms of percent classification accuracy for the six videos of the proposed dataset are given in Figure 3.5, which shows that the proposed ontology tool can assess the risk of collision in the road scenes of the dataset with high accuracy. Results are reported for two hypotheses regarding estimation of the pedestrian's position with respect to the road. Figure 3.4 explains both hypotheses. The first hypothesis takes into account the centre of the pedestrian's bounding box, and the second takes into account 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 the dataset.

Representative examples of risk assessment from the dataset are presented in Figure 3.6. Two of those examples are described in detail here. In Figure 3.6a, the object of interest is the pedestrian and the object's location is the road. The pedestrian's speed is estimated at 2.6 pixels per frame. According to Eq. 3.24, this speed is classified as low. The object's direction is 90 degrees with respect to the car driver's perspective. The ontology tool infers that the situation poses a high level of risk. The key feature that influences the decision is *pedestrian location*. In the example illustrated in Figure 3.6b, the pedestrian's speed is 5.9 pixels per frame, which, according to the Eq. 3.24, is classified as medium speed. The object's direction is 90 degrees with respect to the car driver's perspective. The ontology tool infers that this scene does not pose risk.

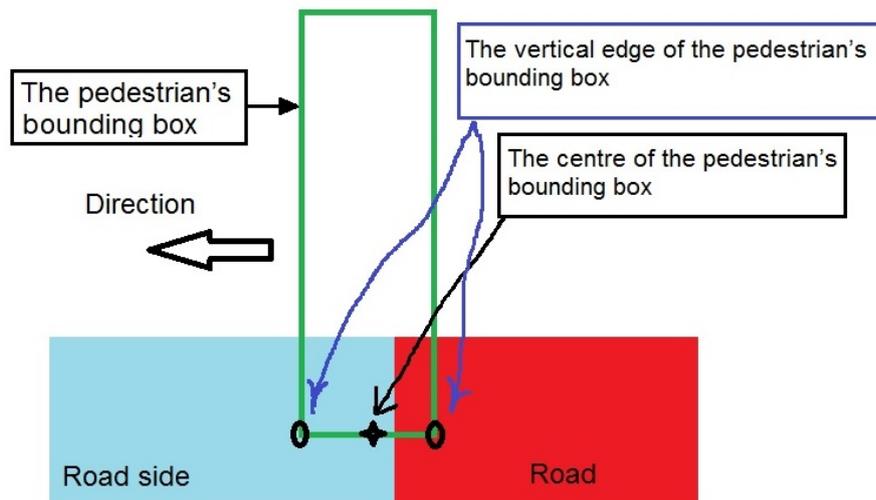


Figure 3.4: An explanation of the two hypotheses for estimating pedestrian position, the *vertical-edges* hypothesis and the *centre-of-bounding-box* hypothesis.

Table 3.1: Summary of the proposed dataset for pedestrian behaviour in road scenes

Sequences	Num. frames
Case 1 footage 1	83
Case 1 footage 2	75
Case 1 footage 3	59
Case 1 footage 4	105
Case 1 footage 5	64
Case 1 footage 6	131
All frames	517

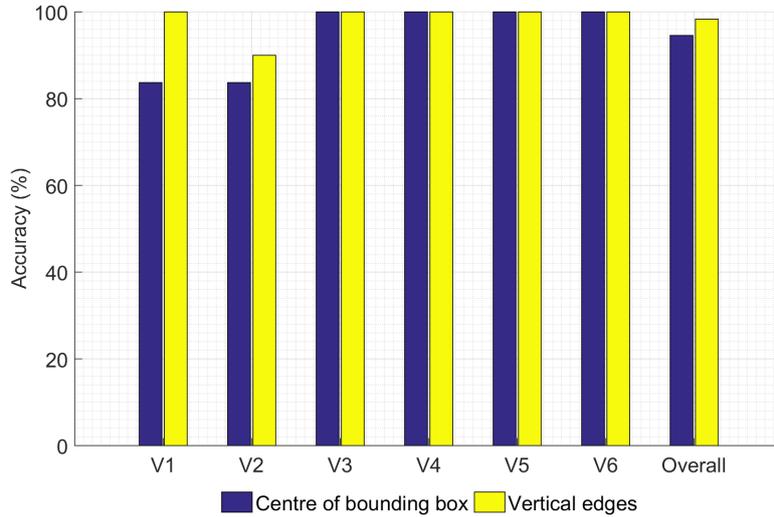


Figure 3.5: Experimental risk-assessment accuracy for the pedestrian portion of the structure of the proposed ontology.

In Figure 3.7, the output of the ontology’s inference tool over time is plotted. The output is obtained from a video from the proposed dataset together with the extracted features. There are four key events in this video, each of which is explained individually.

- At  $K_0$ , the pedestrian ( $P$ ) is waiting on the roadside ( $rs$ ) with no speed and no direction. Therefore, according to the rules pro-

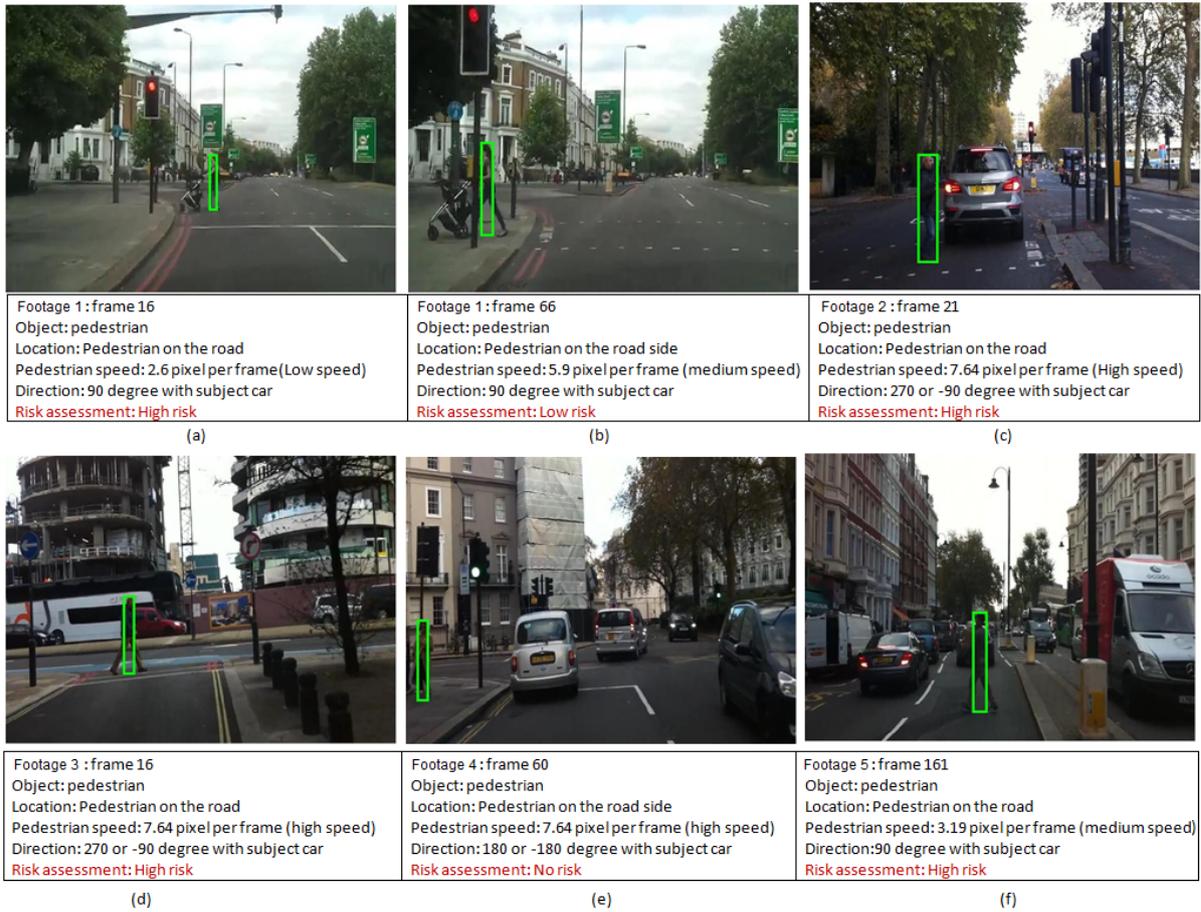


Figure 3.6: Risk-assessment examples.

posed in Section 3.1.3, the ontology tool inferred that the situation does not conceal risk.

- At  $K_1$ , the pedestrian ( $P$ ) on the roadside ( $rs$ ) has started walking with low speed towards the road ( $r$ ). Therefore, according to the rules proposed in Section 3.1.3, the ontology tool inferred that the situation poses a medium level of risk.
- At  $K_2$ , the pedestrian ( $P$ ) on the roadside ( $rs$ ) is walking with high speed towards the road ( $r$ ). Therefore, according to the rules proposed in Section 3.1.3, the ontology tool inferred that the situation poses a 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 rules proposed in Section 3.1.3, the ontology tool inferred that the situation poses a high level of risk.

It can be seen that for the key events  $K_3$  and  $K_4$ , the ontology's reasoning tool inferred the same level of risk, even though the pedestrian's speed is different in each event. This is due to an important property that appears in both events, that 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 key feature between the two events is the direction of the pedestrian *towards the road*, regardless of the pedestrian's speed. By contrast, the role of *speed* is more important when comparing events  $K_0$  and  $K_1$ , because in those events it is the only factor that influences the output of the ontology tool.

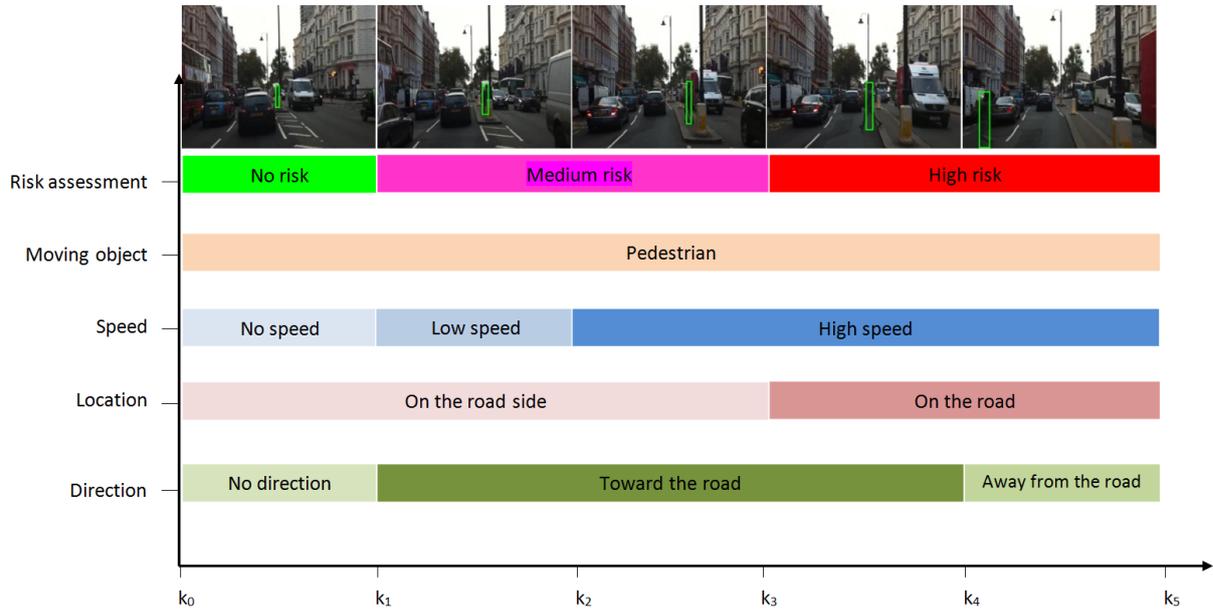


Figure 3.7: Event-based graphical illustration of the ontology results.

### 3.3 Summary

This chapter proposes a novel ontology that tackles the problem of automatic risk assessment in unpredictable road traffic environments. A framework for video-based assessment of the degree of risk of collision in a road scene encompassing the ontology is also presented in this chapter. Unlike previous work in SA, in which 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, the ontology does not assume that road users obey the traffic rules; 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 perfor-

mance 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.

In this chapter, pedestrians, speed and location are detected manually. The following chapters investigate the selection of the appropriate approaches from computer vision methods for fully automatic pedestrian, speed and location detection and tracking. Thus, the framework tends to be a fully automatic risk assessment.

# A NEW METHOD FOR EVALUATION OF VIDEO SEGMENTATION QUALITY

The framework proposed in this thesis uses as input video captured by a single monocular video camera. Because this input consists of high-dimensional data, it is necessary to reduce the data's dimension. As explained in Section 2.2, segmentation is a process that can effectively read and interpret the semantic data of this high-dimensional content [Ngan and Li, 2011]. It does so by subdividing images into meaningful segments [Dey et al., 2010, Morris et al., 1986]. Therefore, segmentation is an important dimension-reduction process for high-dimensional video data, such that it becomes a crucial stage in image/video analysis and understanding. Many approaches have been proposed to provide a high quality segmentation; hence, their evaluation also plays an important role in the assessment of the quality of segmentation results. Although, as outlined in Chapter 2, a considerable amount of research has investigated video segmentation, little research has focused specifically on the evaluation of video segmentation quality.

As stated in Section 2.2.2, several criteria for the evaluation of image segmentation have been proposed. Due to the differences between image segmentation and video segmentation, it is necessary to propose new criteria by considering additional characteristics of high quality video segmentation.

Here, this study's findings make several contributions to the current state of the art. First, new criteria for high quality video segmentation are proposed, that consider the additional characteristics of the stability of the boundaries and consistent region identity between consequent frames. Second, on the basis of the new criteria, an online method for the evaluation of video segmentation quality is proposed; the method can be used both for supervised and unsupervised evaluation. Third, a synthetic video set is designed to evaluate the evaluation methods of video segmentation with this video set.

This chapter is organised as follows. In Section 4.1, the proposed criteria and metrics are discussed. In Section 4.2, a detailed overview of the proposed method is given. Section 4.3 provides the evaluation and results. The main conclusions of the chapter are summarised in Section 4.4.

## **4.1 Proposed criteria and metrics**

An initial step for any evaluation process is determining the criteria of evaluation. As discussed in Section 2.2.2, researchers have proposed several criteria for the evaluation of image segmentation. Due to the differences between image segmentation and video segmentation, there are no established criteria for evaluating the quality of video segmentation. It is, therefore, crucial to propose new criteria by considering

additional characteristics related to high quality video segmentation.

For evaluating video segmentation quality, in addition to taking account of the proposed criteria for the evaluation of image segmentation, the stability of the boundaries and consistent region identity between consequent frames should be evaluated [Grundmann et al., 2010]. Given this concern, and in light of the criteria proposed by Haralick and Shapiro, the following set of criteria is proposed:

1. The regions must be uniform, homogeneous, simple and without holes.
2. Adjacent semantic regions should have significant differences with respect to the characteristic on which they are uniform.
3. Corresponding regions between consequent frames should be consistent.
4. Boundaries of the segmented frame should be smooth, stable and accurate when compared with the boundaries of the original frame.

All of these criteria are applicable to evaluating the quality of video segmentation in a supervised way, whereas the method proposed in this chapter offers evaluation in both a supervised and unsupervised way. The unsupervised-evaluation component of this study is based on low-level image features, in accordance with previous unsupervised methods [Zhang et al., 2008]. For this reason, the second criterion will not be used when evaluating the quality of video segmentation, because it is difficult to find meaningful adjacent segments without semantic information. In the next section, the metrics for measuring the

Table 4.1: Summarise the applied criteria, measures and metrics

Criteria	Measures	Metrics
No_1	Intraregion uniformity and homogeneity	$F_{RC}$ [Rosenberger and Chehdi, 2000] and $Tex\_var$ [Correia and Pereira, 2003]
No_3	Temporal-region consistency	Pearson's correlation and GCI [Martin et al., 2001]
No_4	Boundary stability and accuracy	F-measure [Martin et al., 2001]

quality of video segmentation according to the remaining three criteria are considered (Table 4.1): those for measuring intraregion uniformity and homogeneity (criterion 1), those for measuring region consistency between consequent frames (criterion 3) and those for measuring boundary accuracy (criterion 4).

#### 4.1.1 Intraregion uniformity and homogeneity

The uniformity of regions can be divided into two categories, colour uniformity and texture uniformity. The former means that the pixel colours of a region should have similar values; the latter means that each region should have consistent texture. In [Zhang et al., 2008], the intraregion uniformity metrics are classified into four classes, based on colour error, squared colour error, texture and entropy. Two simple and easy-to-understand metrics are selected:  $F_{RC}$  [Rosenberger and Chehdi, 2000], which measures the intraregion colour disparity and is based on squared colour error, and texture variance  $Tex\_var$  [Correia and Pereira, 2003], which measures texture uniformity and is based on the variance of the Y, U and V layers.

### 4.1.2 Temporal-region consistency

A number of methods can be used to evaluate consistency between two regions. In addition, a number of metrics have been designed specifically to measure the similarity between two ground-truth images, such as variation of information (VI) [Meilă, 2003, Unnikrishnan et al., 2005], global consistency error (GCE) [Martin et al., 2001] and probabilistic Rand index (PRI) [Unnikrishnan et al., 2005]. GCE and VI are designed to compare two segmentations, whereas PRI is designed to compare more than two segmentations. VI is an information based metric, which considers mutual information between two segmentations, whereas GCE is a region-based metric, which is designed to quantify the consistency between image segmentations of different granularities [Unnikrishnan et al., 2007].

In video segmentation, corresponding regions from consecutive frames should have consistent colour and granularity. Here, a combination of GCE and positive correlation is proposed for the evaluation of the consistency between two consecutive frames. Although GCE has been used for image-segmentation evaluation, it has not been combined with positive correlation. This combination has the advantage of taking into account both the consistency of region granularity and the colour consistency of regions between consecutive frames.

### 4.1.3 Boundary stability and accuracy

Boundaries can be defined as edges that separate two regions. The main purpose of boundary detection is to characterise semantic objects in the scene by drawing a borderline between adjacent semantic regions, without considering their interior features [Hoogs and Collins,

2006]. To measure the accuracy between the boundaries of segmented and original frames, the F-measure metric [Martin et al., 2001] is used, which is the most popular boundary-based metric for the evaluation of image segmentation [Galasso et al., 2013]. A method for detecting the boundaries of original frames is developed in the case of unsupervised evaluation, and ground-truth boundaries are used in the case of supervised evaluation. The proposed method will be explained in the next section.

## 4.2 Proposed evaluation method

In this chapter, an evaluation method based on the new criteria is proposed. The method can be used for both supervised and unsupervised evaluation. The former uses ground-truth boundaries; the latter uses the boundaries detected in the original frame, for which a combination of low-pass filtering to remove noise and multiscale edge detection is used.

The proposed evaluation method uses the detected boundaries twice. First, they are used to handle and specify the regions of the segmented frame. Then, as outlined in Section 4.2.1, their intraregion uniformity is measured. Second, their accuracy is evaluated by comparing boundaries of the segmented frame with the ground truth boundaries of the same frame.

### 4.2.1 Intraregion uniformity

Selecting the semantic regions that compose an image requires either human assistance or a ground-truth template, neither of which are available for unsupervised segmentation. It is possible to overcome this

problem by detecting and using the boundaries of the original video frames. Thus, the process of evaluating intraregion uniformity consists of the following three steps:

1. *Detecting boundaries.* For supervised evaluation, the method uses the ground-truth boundaries. By contrast, to detect boundaries in the case of unsupervised evaluation, a method relying on a combination of low-pass filtering to remove noise and multiscale edge detection is used. An example of this form of boundary detection is shown in Figure 4.1.
2. *Selecting regions from the segmented frame.* The detected boundaries produced from the previous step are used to select the regions of the segmented frame. Then, quad-tree image decomposition is used to separate the segmented frame into a number of rectangular areas that do not contain any boundaries from the original video frame. The uniformity of each of the rectangular areas is evaluated in the next step.
3. *Evaluating intraregion uniformity.* The selected regions produced from the previous step are evaluated using two metrics selected in Section 4.1.1, namely  $F_{RC}$  and  $Tex\_var$ .

The metrics  $F_{RC}$  and  $Tex\_var$  will be explained individually. Let  $N$  be the total number of regions of segmented image  $I$ , with height  $I_x$  and width  $I_y$ ,  $j$  be the index of regions  $j \in (1, 2, 3, \dots, N)$ ,  $R_j$  represent set of pixels in the region  $j$  where  $R_j \subset (\cup_{j=1}^N (R_j))$ ,  $S_j$  be the area of region  $j$ ,  $C_x(P)$  be the colour intensity value for pixel  $P$  ( $x \in$  red, green, or blue component) and the area of the full image be  $S_I = I_x \times I_y$ .

The mean value of component  $x$  in region  $j$  can be defined as follows:

$$\widehat{C}_x(R_j) = \frac{1}{S_I} \sum_{P \in R_j} C_x(P), \quad (4.1)$$

$F_{RC}$  is based on the squared colour error and measures the intraregion colour disparity. Squared colour error can be defined as follows:

$$e_x^2(R_j) = \sum_{P \in R_j} (C_x(P) - \widehat{C}_x(R_j))^2, \quad (4.2)$$

The first metric,  $F_{RC}$  can be defined as follows:

$$D(I) = \frac{1}{N} \sum_{j=1}^N \frac{S_j}{S_I} \times e_x^2(R_j), \quad (4.3)$$

where  $D(I)$  is the  $F_{RC}$  colour disparity of image/frame  $I$ , and  $e_x^2(R_j)$  is the squared colour error of region  $R_j$ .

The second metric, texture variance  $Tex\_var$  [Correia and Pereira, 2003] is defined as follows:

$$Tex\_var(I) = \frac{1}{N} \sum_{j=1}^N \frac{1}{5} \left( 3 \times \sigma_{y(R_j)}^2 + \sigma_{u(R_j)}^2 + \sigma_{v(R_j)}^2 \right), \quad (4.4)$$

where  $Tex\_var(R_j)$  is the texture variance of the region( $R_j$ ) and  $\sigma_y$ ,  $\sigma_u$  and  $\sigma_v$  are the variances of the  $Y$ ,  $U$  and  $V$  components in region  $R_j$ , respectively.

Both  $D(I)$  and  $Tex\_var(I)$  metrics are normalised to intraregion uniformity  $I_U$ , and texture uniformity  $T_U$ , respectively, by the following

function:

$$\eta = \left( \frac{1}{1 + \frac{\nu^{0.5}}{128}} - 0.5 \right) \times 2, \quad (4.5)$$

where  $\eta$  and  $\nu$  represent the normalised value (between 0 and 1) and the initial value (between 0 and  $128^2$ ) of the metrics, respectively.

A real scene can consist of both colour and texture regions, and it is difficult to determine which region category is predominant in the scene. For this reason, both colour uniformity and texture uniformity to measure the region uniformity are used, and their averages are calculated to take them both into account.



Figure 4.1: Example of unsupervised boundary detection for frame number 1, from left to right: Soccer sequence and Ice sequence.

### 4.2.2 Region consistency

The content of consecutive frames in video sequences is usually not completely identical, but there is consistency and similarity between them, the degree of which is dependent on the complexity of the sequence. As discussed previously, according to the criteria listed in Section 4.1, identical regions between consequent frames should be consistent in terms of both colour and granularity. The main purpose of this section is to ensure this consistency according to both metrics. To do so, the minimum value between the global consistency index GCI used to evaluate granularity consistency and the positive correlation used to evaluate colour consistency is employed. GCI can be explained as follows.

Let  $\setminus$  denote set difference and  $|x|$  the cardinality of set  $x$ . Let  $S_1$  and  $S_2$  be two segmentations. For a given pixel  $p_i$ , consider the segments that contain  $p_i$  in  $S_1$  and  $S_2$ . Let these sets of pixels be denoted by  $R(S_1, p_i)$  and  $R(S_2, p_i)$ , respectively. The local refinement error is defined as follows:

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) \setminus R(S_2, p_i)|}{|R(S_1, p_i)|}, \quad (4.6)$$

$$GCE(S_1, S_2) = \frac{1}{n} \min \left( \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right), \quad (4.7)$$

$GCE(S_1, S_2)$  is the global consistency error between frames  $S_1$  and  $S_2$ , and  $n$  is the number of pixels.

$$GCI(S_1, S_2) = 1 - (GCE(S_1, S_2)), \quad (4.8)$$

In this study, the GCI is applied to each frame, which means that  $S_1$  and  $S_2$  represent two consecutive frames.

The positive correlation between consecutive frames can be calculated as follows:

$$Corr(S_1, S_2) = \begin{cases} r(S_1, S_2) & r(S_1, S_2) \geq 0, \\ 0 & r(S_1, S_2) < 0, \end{cases} \quad (4.9)$$

where  $r(S_1, S_2)$  is the Pearson's correlation between frames  $S_1$  and  $S_2$  and can be defined as follows:

$$r(S_1, S_2) = \frac{n \sum (S_1 S_2) - (\sum S_1)(\sum S_2)}{\sqrt{(n \sum S_1^2 - (\sum S_1)^2)(n \sum S_2^2 - (\sum S_2)^2)}}, \quad (4.10)$$

### 4.2.3 Boundary assessment

F-measure [Martin et al., 2001] is the most popular metric in this area, as discussed in Section 4.1. Let  $F$  denote the F-measure.

$$F = \frac{2 * P * R}{P + R}, \quad (4.11)$$

where  $P$  is the precision of the boundaries and  $R$  is the recall of the boundaries.

### 4.2.4 Combining metrics

The selected metrics explained in the previous sections are combined as a version of the F formula.  $F$  is the harmonic mean of precision and recall, with precision penalising oversegmentation and recall penalising undersegmentation, both of which are important for evaluating the

quality of video segmentation. In this study, the precision  $P$  is updated to  $P'$  to include the metrics evaluating region uniformity and consistency, which also play an important role in penalising over- and undersegmentation.

$$F' = \frac{2 * P' * R}{P' + R}, \quad (4.12)$$

$$P' = \frac{P + \alpha}{2}, \quad (4.13)$$

where  $P'$  is the updated precision and the average between precision  $P$  and  $\alpha$ .  $\alpha$  can be defined as follows:

$$\alpha = \frac{2 * U * C}{U + C}, \quad (4.14)$$

where  $\alpha$  is the harmonic mean between intraregion uniformity  $U$  and consistency  $C$ . Both  $U$  and  $C$  can be defined as follows:

$$U = \frac{I_U + T_U}{2}, \quad (4.15)$$

$$C = \min(GCI, Corr), \quad (4.16)$$

where  $I_U$  is the minimum value of normalised intraregion uniformity among R, G and B layers, obtained using Eqs. 4.3 and 4.5,  $T_U$  is the normalised texture uniformity obtained using Eqs. 4.4 and 4.5,  $GCI$  is the minimum value of the global consistency index among R, G and B layers (Eq. 4.8) and  $Corr$  is the minimum value of positive correlation among R, G and B (Eq. 4.9).

### 4.3 Evaluation and results

The proposed method is evaluated on both synthetic and real video data. The synthetic videos are created with representations of different types of segmentation defects. The real videos are selected from a publically available dataset [Chen and Corso, 2010]. The details of each dataset will be explained in a later section. The proposed method evaluates video segmentation quality better than the state-of-the-art method, on different types of content. The results of each dataset are shown below.

#### 4.3.1 Synthetic data

Two synthetic videos with a length of 100 frames each are created. The first video depicts four differently coloured circles moving from different corners towards each other, meeting in the middle and then moving to the opposite corners (Figure 4.3). The second video represents different defects in the segmentation of the first video, such as over- and undersegmentation, undetected objects, inconsistent object identity (swapping of identity between objects), etc. Moreover, the ‘correctly segmented’ frames between the ‘defective segmentations’ are inserted to represent inconsistent temporal segmentations (Figure 4.2).

#### 4.3.2 Real video data

The real video dataset is from [Chen and Corso, 2010] and is a subset of the *Xiph.org* videos. The selected dataset used in this study can be divided into three groups: ground truth, oversegmented and undersegmented. Six different videos labelled with 24-class semantic pixel

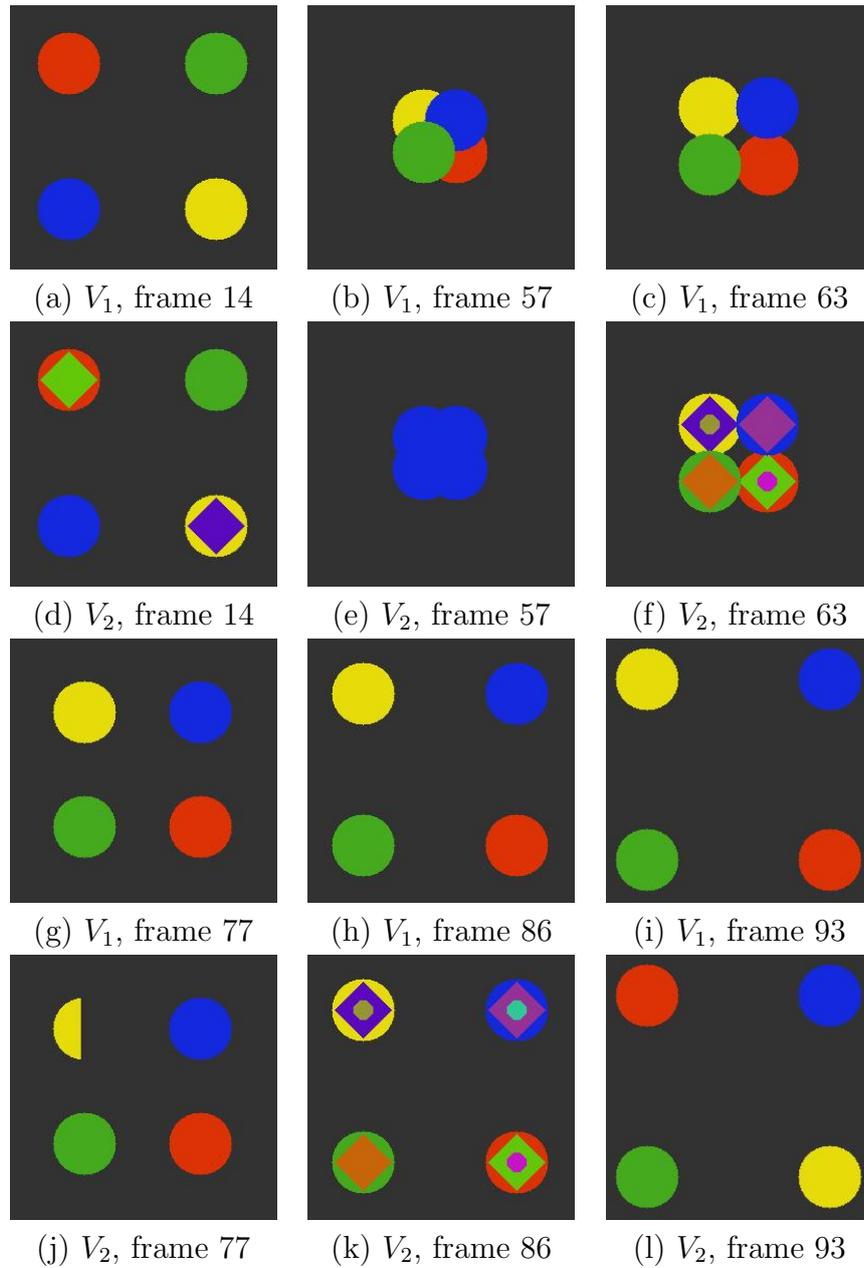


Figure 4.2: Some examples of various defects in second video  $V_2$ : the first frame group presented in (a), (b), (c), (g), (h) and (i) represent the correctly segmented frames in the first video  $V_1$ , but the second group presented in (d), (e), (f), (j), (k) and (l) represent the defective segmentation in the second video  $V_2$ .

labelling are used as ground truth [Chen and Corso, 2010]. For each video, three degrees of undersegmentation are created from ground-truth frames, and three degrees of oversegmentation are created using

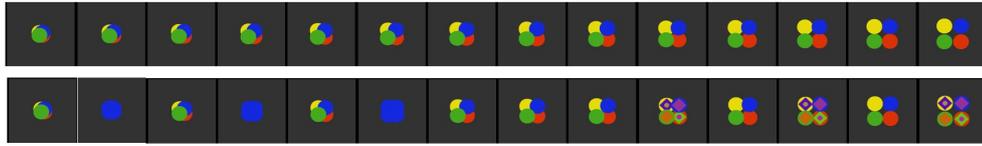


Figure 4.3: Sample of the synthetic video; first row is the segmented frames in first video  $V_1$  and the second is the defective segmentation in second video  $V_2$ , both sequences are in the same order.

the hierarchical graph-based method [Grundmann et al., 2010]. The length of the videos varies from 69 to 86 frames. An example of this real video dataset is shown in Figure 4.4.

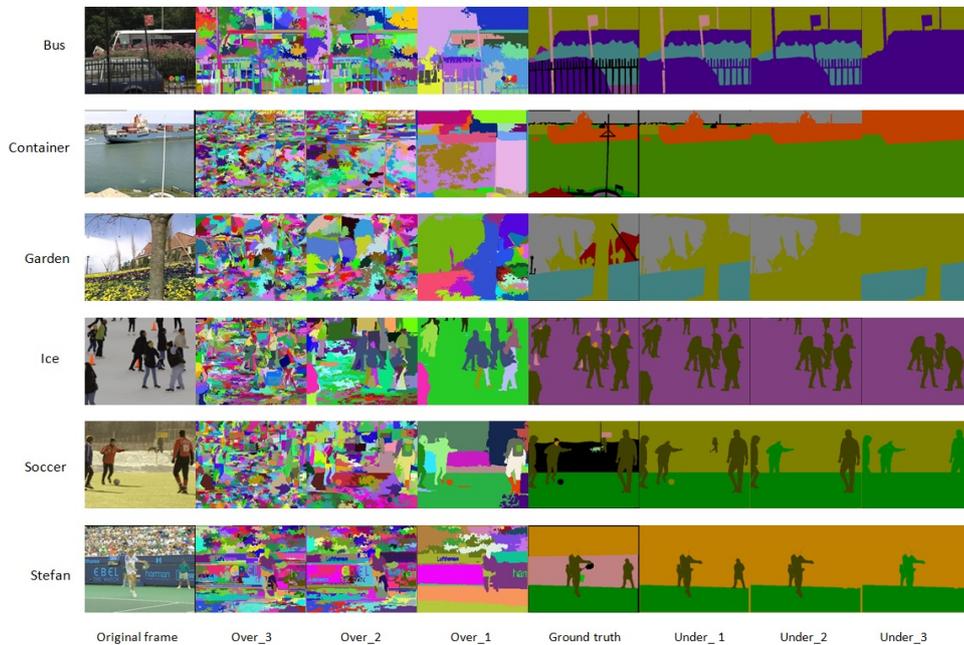
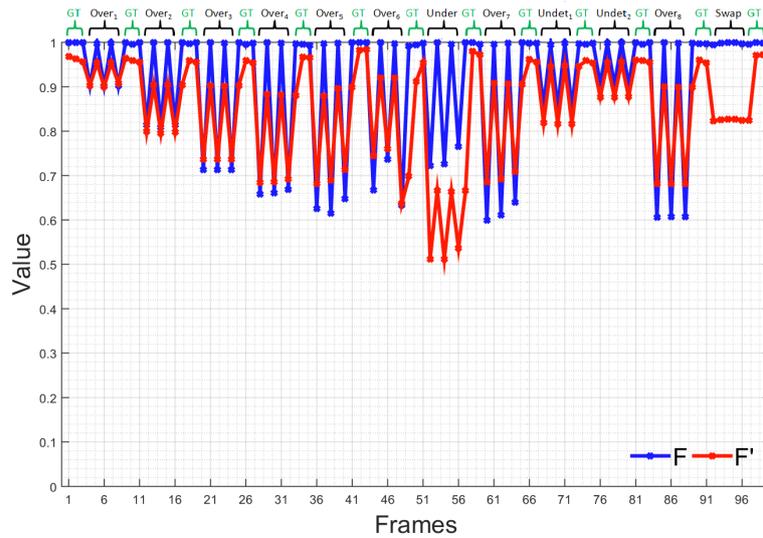


Figure 4.4: Visual comparison of the six aspects of segmentation quality and ground truth.

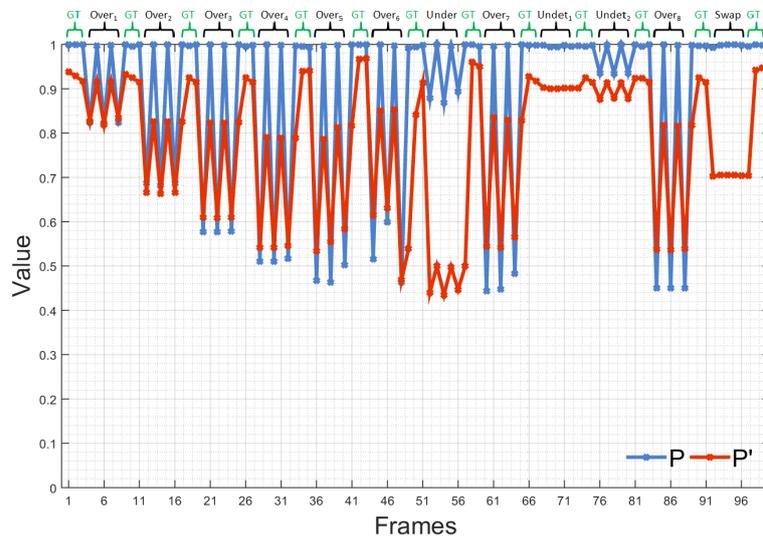
### 4.3.3 Results on synthetic video

This example explains the ability of  $F$  and  $F'$  to evaluate different types of segmentation defects. Both  $F$  and  $F'$  are applied to the synthetic video dataset described in the previous section. The results of  $F$  are accurate in most of the cases, but it is not as strict as  $F'$  in penalising

inconsistent object identity and undersegmentation. Figure 4.5 shows the differences between  $F$  and  $F'$  and between  $P$  and  $P'$ . Frames 53 to 57 are undersegmented, and frames 94 to 97 contain inconsistent object identity.



(a)



(b)

Figure 4.5: Result of: (a)  $F$  and  $F'$ ; (b)  $P$  and  $P'$ .

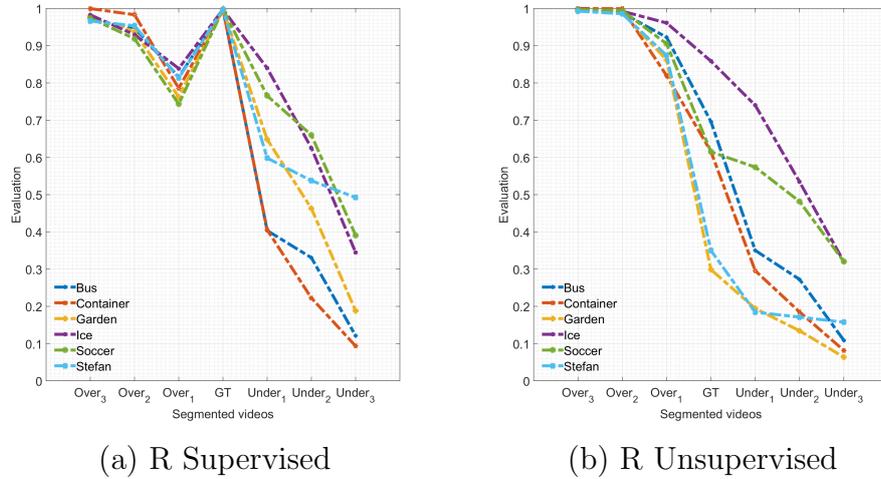
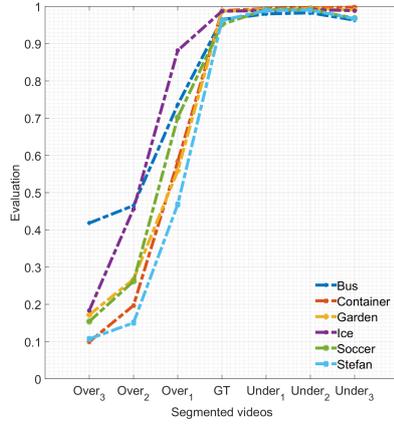


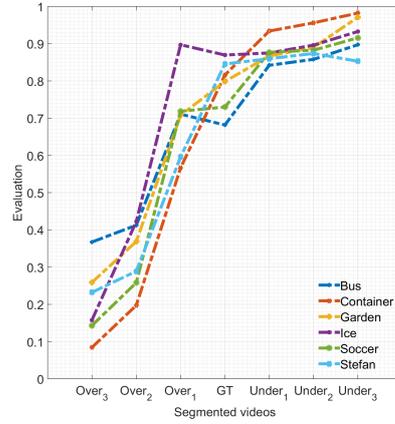
Figure 4.6: Segmentation evaluation results using recall R for the supervised and unsupervised cases, for six video sequences (Bus, Container, Garden, Ice, Soccer and Stefan). Each video has seven degrees of segmentation, ground truth, three degrees of undersegmentation and three degrees of oversegmentation.

#### 4.3.4 Results on real video

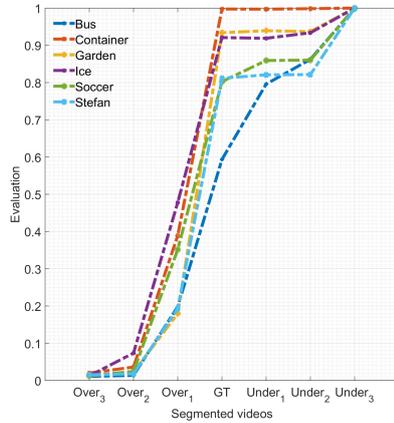
In addition to the synthetic video evaluation, the proposed method is evaluated on the six real videos, as outlined in Section 4.3.2. For each real video from the dataset, seven segmentations of different quality, containing ground truth, three degrees of oversegmentation, and three degrees of undersegmentation are created (Figure 4.4). Figures 4.9 and 4.10 present the comparative information on  $F$  and  $F'$  over these segmentations. Although  $F$  and  $F'$  are approximately the same for the undersegmented and ground-truth segmentations, their behaviour for the oversegmented areas is different.  $F'$  is more consistent with the perceptual quality of the segmentations  $Over\_3$ ,  $Over\_2$  and  $Over\_1$  than  $F$ , where a significant difference can be observed in the quality of  $Over\_2$  and  $Over\_1$  (Figures 4.6, 4.7, 4.8, 4.9 and 4.4). These differences are more clear when comparing  $P$  with  $P'$  for both supervised and unsupervised cases (Figure 4.8). This is due to the effect of the metrics



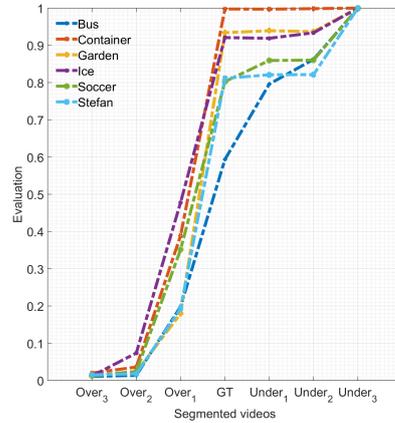
(a) U Supervised



(b) U Unsupervised



(c) C Supervised

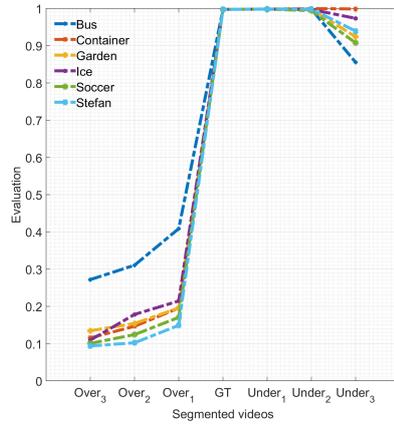


(d) C Unsupervised

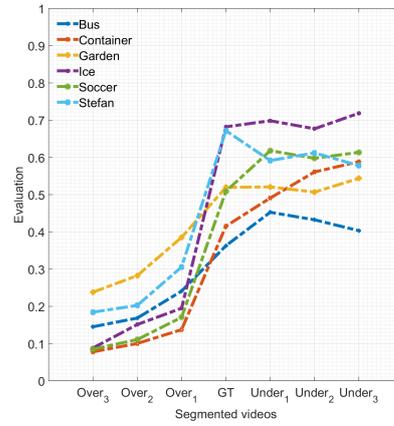
Figure 4.7: Segmentation-evaluation results using intraregion uniformity U and temporal-region consistency C for the supervised and unsupervised cases, for six video sequences (Bus, Container, Garden, Ice, Soccer and Stefan). Each video has seven degrees of segmentation, ground truth, three degrees of undersegmentation and three degrees of oversegmentation.

of intraregion uniformity and temporal-region consistency (Figure 4.7).

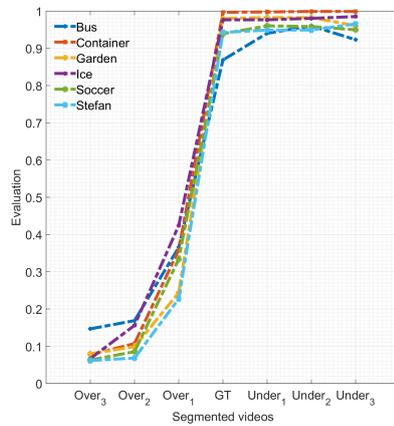
In addition, the unsupervised boundary recall shows promising results compared with the supervised boundary recall. Specifically, a notable difference can be perceived in evaluating *Over\_2* and *Over\_1*; the unsupervised boundary recall shows that the perceptual quality of *Over\_1* is better than *Over\_2*, which is true, while it receives a lower score than



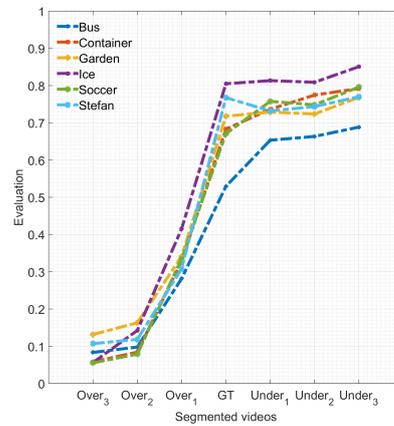
(a) P Supervised



(b) P Unsupervised



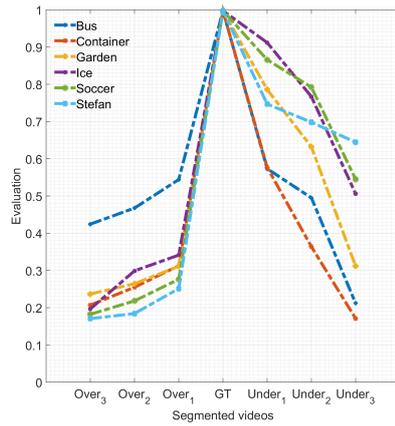
(c) P' Supervised



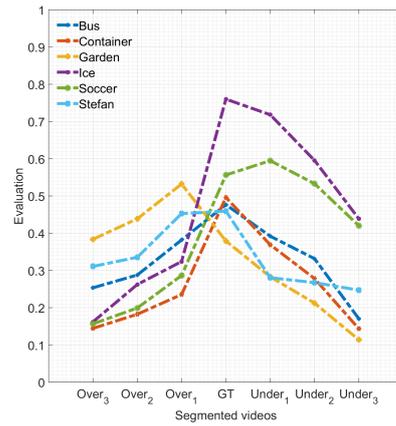
(d) P' Unsupervised

Figure 4.8: Segmentation evaluation results using P and P' for the supervised and unsupervised cases, for six video sequences (Bus, Container, Garden, Ice, Soccer and Stefan). Each video has seven degrees of segmentation, ground truth, three degrees of undersegmentation and three degrees of oversegmentation.

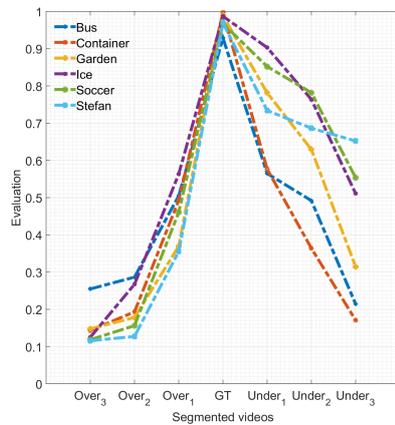
*Over\_2* by the supervised boundary recall.



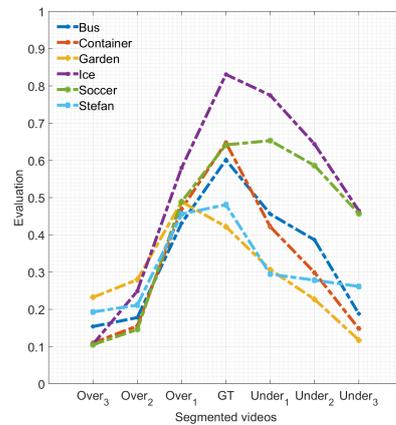
(a) F Supervised



(b) F Unsupervised



(c) F' Supervised



(d) F' Unsupervised

Figure 4.9: Segmentation evaluation results using F and F' for the supervised and unsupervised cases, for six video sequences (Bus, Container, Garden, Ice, Soccer and Stefan). Each video has seven degrees of segmentation, ground truth, three degrees of undersegmentation and three degrees of oversegmentation.

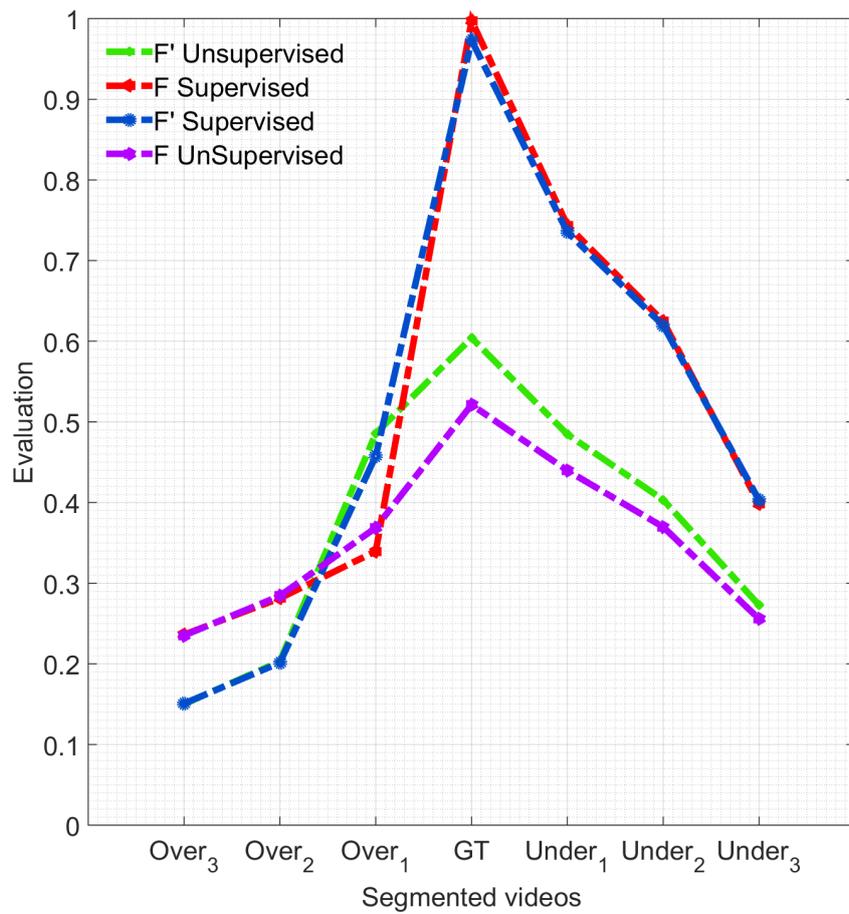


Figure 4.10: Results of real video evaluation using F and F' evaluation metrics, for the supervised and unsupervised case. The average score of all videos for each segmentation quality are reported.

## 4.4 Summary

Evaluation of video segmentation quality is an important process to assess the quality of segmentation results. As the initial step of the evaluation process, this chapter proposes new criteria of high quality video segmentation. The proposed criteria consider additional characteristics of high quality video segmentation: the stability of the boundaries and consistent region identity between consequent frames.

Then, on the basis of these new criteria, the chapter proposes an online method for the evaluation of video segmentation quality that takes into account the characteristics of both boundaries and regions.

In addition, a test video set is designed specifically for the purpose of evaluating the performance of the proposed method. The proposed method is evaluated and compared against a supervised state-of-the-art evaluation method in both supervised and unsupervised modes.

The results show that the proposed method can evaluate the quality of video sequences better than F, on different types of content. It can do so because it takes into account region uniformity and consistency between consecutive frames, which is included in the new set of criteria.

# EVOLVING GMMS FOR ROAD-TYPE CLASSIFICATION

As explained in the literature, online road-type classification is crucial in the area of SA and risk assessment, because each road type requires a specific driving behaviour and this is a valuable clue for autonomous vehicles' assessment of upcoming risks. Therefore, road types are included as risk factors in the general structure of the ontology proposed in Chapter 3 . A number of studies have investigated road-environment classification [Mioulet et al., 2013, Tang and Breckon, 2011], In this chapter, a new method for classifying road types on the basis of the data obtained using a monocular camera is proposed. As in [Tang and Breckon, 2011], four classes of problems are considered: motorway, off-road, trunk road, and urban road.

The main contribution of Chapter 5 is a new online vision-based road-type classification method. The proposed method uses video captured by a single video camera, and unlike existing methods, it takes into account the visual information of the whole scene by segmenting the video frames into temporally consistent frame segments. To this

end, a video segmentation algorithm based on evolving Gaussian mixture models (GMMs) is used.

Experimental results on real-world data indicate that the proposed method outperforms the state-of-the-art method in this area in both classification accuracy per road type and overall classification accuracy.

This chapter is organised as follows. In Section 5.1, an overview of evolving GMMs is presented. In Section 5.2, the models of different road types are discussed. In Section 5.3, a detailed description of the classification approach is given. Section 5.4 presents the experimental results. The main conclusions of the chapter are summarised in Section 5.5.

## 5.1 Online video segmentation

The road-type models are built using the evolving GMM algorithm from [Kaloskampis and Hicks, 2014]. In this section, a high-level overview of this algorithm is given and its use is justified. As mentioned in Chapter 2, state-of-the-art methods use three subregions as regions of interest for the driving environment, namely, road, road edge and roadside. The features of these three subregions are used as the key information during classification. However, there is no guarantee that the subregions will capture all the key information. In certain cases, in fact, the subregions are unlikely to contain the key information, such as when the car turns left or right, or if it is driven on a rough road. To tackle this issue, the method proposed in this chapter uses the information from all regions within a video frame. At the same time, because the road-classification method is considered to work in real time, it should handle this information efficiently. For these reasons,

the evolving GMM algorithm from [Kaloskampis and Hicks, 2014] is used.

Several videos are used to build the model for each road type. Each of these videos is processed as follows. For every frame in the video, visual features from each of its pixels are extracted. Then, a GMM using the features of the frame is built. After the GMM is built, all the features extracted from the pixels are discarded. Thus, each frame is represented by a GMM rather than its pixel features, which saves a significant amount of computer storage space and memory (in this case study, it is estimated that the GMM representation of a video frame takes up only 0.03% of the memory that its pixel features would require).

The representation of a video sequence could be simply the concatenation of the components of the GMMs that were built on all frames of the sequence. However, this would lead to a complex model that would include a large number of overlapping components. The evolving GMM algorithm from [Kaloskampis and Hicks, 2014] overcomes this problem for the following reasons. After concatenating the GMM components built on all the video frames, this algorithm merges the components using a modified version of the expectation-maximisation algorithm. This process results in a compact, *merged* model with no overlapping components. The size of this merged model is similar to that of a simple GMM generated on a single frame. For more details on the merging process, see [Kaloskampis and Hicks, 2014].

The final model for a road type results from the concatenation of all merged models that were built on video sequences illustrating that road type. To segment a video frame, each pixel in the frame is attributed

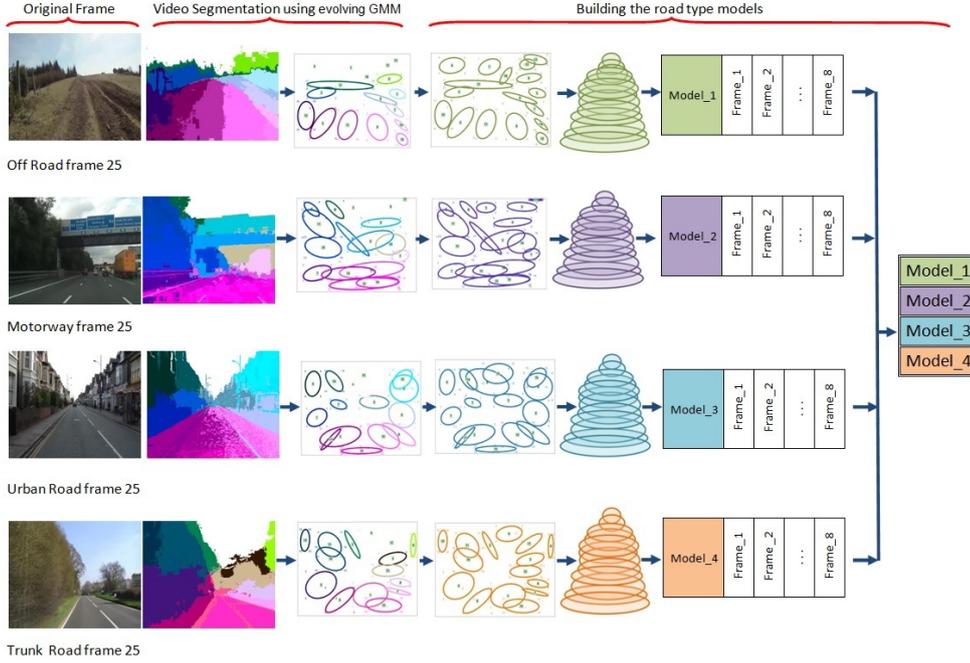


Figure 5.1: Process of building road-type models.

to a segment according to its probability as estimated with the PDF of the final model. The chosen method is suitable for online applications; moreover, it provides consistent segmentation by preserving long-term information throughout the frames.

## 5.2 Building the road-type model

In this section, the process of building a model for each road type is described. This process is illustrated in Figure 5.1. For each road type,  $i$ , a set  $S_i$  of  $m$  image sequences illustrating road type  $i$  is selected. The set  $S_i$  is given by:

$$S_i = \{I_i^{(1)}, I_i^{(2)}, \dots, I_i^{(m)}\}, \quad (5.1)$$

where  $I_i^{(n)}$ ,  $n \in \{1, 2, \dots, m\}$  is an image sequence of road type  $i$ .

Then, visual features from every frame of each image sequence in

set  $S_i$  are extracted. Following [Goldberger and Greenspan, 2006b, Kaloskampis and Hicks, 2014] this is achieved by representing each pixel in each frame with a five-dimensional vector that includes the pixel's colour descriptor in the *Lab* colour space and the pixel's spatial coordinates. If  $F_i^{(n)}$  denotes the feature representation of an image sequence  $I_i^{(n)}$ , the set of feature representations  $S_i'$  can be obtained as follows:

$$S_i' = \{F_i^{(1)}, F_i^{(2)}, \dots, F_i^{(m)}\}, \quad (5.2)$$

The evolving GMM algorithm from [Kaloskampis and Hicks, 2014] is then applied to all the feature representations of  $S_i'$ ; thus, for each frame in the set sequences  $S_i$ , each homogeneous region becomes a GMM, by grouping pixels on the basis of feature similarities of their selected five-dimensional feature space  $S_i'$ . Next, all GMMs are labelled manually according to road-type model  $M_i$  categories. Finally, all resulting GMMs illustrating the same road type are concatenated into a unified model. The model  $M_i$ , for road type  $i$ , is given by:

$$M_i = \{L_{ik}\}_{k \in \{1, 2, \dots, N_i\}}, \quad (5.3)$$

where  $L_{ik}$  is the  $k^{th}$  Gaussian in  $M_i$ , and  $N_i$  is the total number of Gaussians in  $M_i$ . In this study, following the suggestion of [Tang and Breckon, 2011], four road types are considered: off-road, motorway, urban road and trunk road.

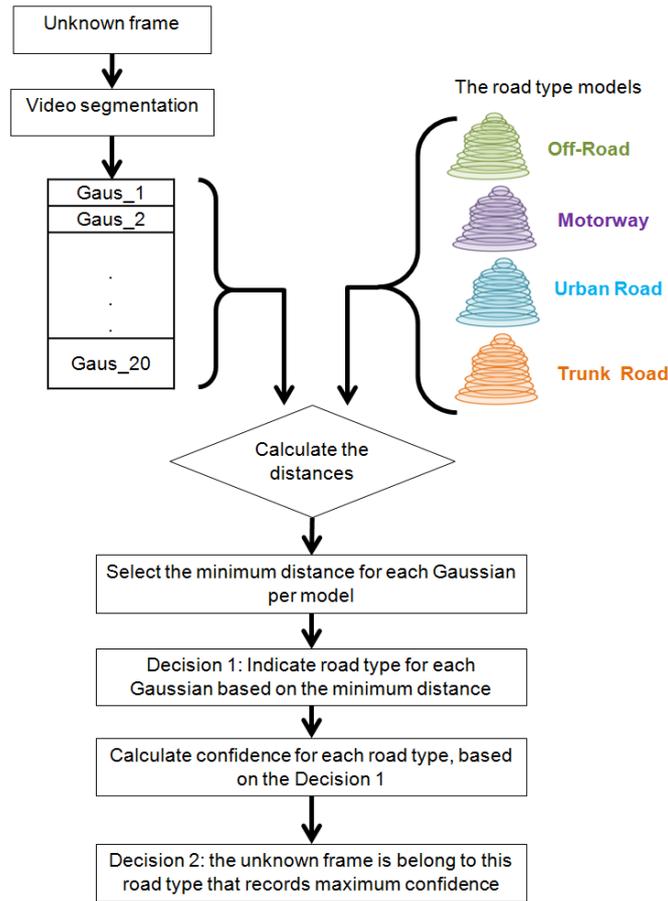


Figure 5.2: Pipeline of the classification process.

### 5.3 Classification

The pipeline of the classification process is illustrated in Figure 5.2. An input frame  $f$  is assigned to a road-type  $M_i$  by estimating its proximity to each road type model. First, Gaussian mixture models  $M_f$  for a given frame are built using the evolving GMM algorithm, which is a GMM estimated on frame  $f$  and given by the following equation:

$$M_f = \{U_{fj}\}_{j \in \{1,2,\dots,N_f\}}, \quad (5.4)$$

where  $U_{fj}$  is the  $j^{th}$  Gaussian and  $N_f$  is the total number of Gaussians in model  $M_f$ .

The next step is to estimate the divergence distance between each Gaussian  $U_f$  from the segmented frame  $f$  and the models, using the Bhattacharyya coefficient [Bhattacharyya, 1943, Kailath, 1967], which is used in many research areas [Mohammad et al., 2015, Lee and Choi, 2000, Choi and Lee, 2003, Goudail et al., 2004, You et al., 2009, Mak and Barnard, 1996, Reyes-Aldasoro and Bhalerao, 2006]. Bhattacharyya coefficient measures the amount of overlap between two statistical populations. The value of the Bhattacharyya coefficient is always between zero and one; zero indicates no divergence between populations, and one indicates complete isolation between populations. This divergence distance is defined as

$$B(U_{fj}, L_{ik}) = \frac{1}{8} (\mu_{fj} - \mu_{ik})^T \Sigma^{-1} (\mu_{fj} - \mu_{ik}) + \frac{1}{2} \log \left( \frac{\det \Sigma}{\sqrt{\det \Sigma_{fj} \det \Sigma_{ik}}} \right), \quad (5.5)$$

where  $B(U_{fj}, L_{ik})$  is the Bhattacharyya distance between the  $j^{th}$  Gaussian of the GMM of  $f$  and the  $k^{th}$  Gaussian of the model  $M_i$ .  $(\mu_{fj}, \Sigma_{fj})$  and  $(\mu_{ik}, \Sigma_{ik})$  are the means and covariances of the  $j^{th}$  Gaussian in  $f$  and the  $k^{th}$  Gaussian in the model  $M_i$ , respectively. For  $\Sigma$ , it is  $\Sigma = \frac{\Sigma_{fj} + \Sigma_{ik}}{2}$ .

The minimum distance between the  $U_{fj}$  and the Gaussians in the model  $M_i$  is then calculated. This distance, denoted by  $\beta_{fij}$ , is estimated as

$$\beta_{fij} = \min \{B(U_{fj}, L_{ik})\}, \quad (5.6)$$

The Gaussians are classified on the basis of the divergence distance. In this way, four road types in the road scene are considered, as explained in Section 5.2. The  $U_{fj}$  is classified as the road-type that is the closest to its model; thus, the classification has four possible decision outcomes. The decision is given by the following equation:

$$D_{fj} = \arg \min_i \|\beta_{fij}\|, \quad (5.7)$$

where  $D_{fj}$  is the classification outcome for the  $j^{th}$  Gaussian of  $f$ . Equation 5.7 returns the road type assigned to each Gaussian.

Having classified the Gaussians of  $f$ , the road-type confidence score  $C_{fi}$  for road type  $i$  is estimated. This score can be defined as the percentage of pixels in  $f$  that vote for this road type:

$$C_{fi} = \sum_{j=1}^{N_f} \text{size}(R_{fj}) \times (D_{fj} = i), \quad (5.8)$$

where  $R_{fj}$  is the segmented region in  $f$  that corresponds to the  $j^{th}$  Gaussian of  $U_{fj}$ . The final decision,  $F_f$ , is made by selecting the road type that maximises the confidence score:

$$F_f = \arg \max_i \|C_{fi}\|. \quad (5.9)$$

## 5.4 Experimental results

A model for each road type using eight videos of 25 frames each was built. Thus, each model is built using 200 frames. The videos used for the urban road model were taken from [Brostow et al., 2008], while the videos for the rest of the road types were taken from YouTube. The frame rate was between 25 and 30 fps, and the resolution of all video frames was resized to 640 x 480. All videos were captured from the car driver’s perspective, with legal and safety speed limits for each road type. The detail of the used videos are listed in Table 5.1. For testing, 800 video frames illustrating each road type were used: they were collected similarly to the videos mentioned above. These frames are not used when building the road-type models.

Table 5.1: Dataset summary for road type classification

Sequences	Training frames	Testing frames	Frame rate F/S	Final resolution	Video format
Off-road	200	800	25	640 x 480	.mp4
Motorway	200	800	25	640 x 480	.mp4
Urban road	200	800	30	640 x 480	.mxf
Trunk road	200	800	25	640 x 480	.mp4

Moreover, the state-of-the-art method from [Mioulet et al., 2013] was implemented to benchmark the performance of the proposed method. The same training and testing datasets described above were used. The method uses random forests [Breiman, 2001] for classification; the state-of-the-art method reported the result of 10 trees with the high-

est accuracy. In this experiment, the number of trees in the forest started from (2, 5, 10, 20, etc.) and then gradually increased to 100 trees, which gives the highest classification accuracy. For more than 100 trees, the gain in classification accuracy is insignificant. In addition, the random-forests classifier is applied 10 times and the results are recorded. Then, the mean and standard deviation of the classification accuracy are reported.

The classification results in terms of percent classification accuracy for both methods are presented in Table 5.2 and Figure 5.3. The proposed method achieves higher classification accuracy than the method in [Mioulet et al., 2013] for each road type individually and, consequently, achieves higher overall classification accuracy. The difference between the two methods is more evident in the classification of the off-road environment. The proposed method achieves 96.8% classification accuracy for this road type, whereas the accuracy of the method in [Mioulet et al., 2013] is 61.3%. This is due to the fact that the latter method extracts its features from three predefined subregions in the video frame. However, there is no guarantee that the key information of the scene will always be contained within these regions. Because the proposed method collects features from the entire scene, it is expected that in environments where the scenery is more variable, such as in the off-road case, the proposed method will achieve higher classification accuracy. Visual comparisons between the methods are provided in Figures 5.5, 5.6, 5.7 and 5.8, and the road-class labels are shown in Figure 5.4.

In addition, the confusion matrix for both the proposed method and the state-of-the-art method are shown in Tables 5.3 and 5.4. The con-

fusion matrix of the proposed method has high values on the main diagonal showing a high accuracy for classifying each road type. It shows that 96.875% of off-road types were classified correctly and 3.125% were misclassified as trunk road; at the same time, there was no misclassification with other road types. Moreover, 99.75% of the motorways were classified correctly with 0.25% misclassified as urban road and no misclassification with other road types. In addition, 100% of the urban types were classified correctly. Finally, 92.25% of trunk roads were classified correctly with 2.625% misclassified as urban road and 5.125% misclassified as motorway. In contrast, the confusion matrix of the state-of-the-art method shows smaller values on the main diagonal, as well as shows a wider range of misclassification between the road type classes.

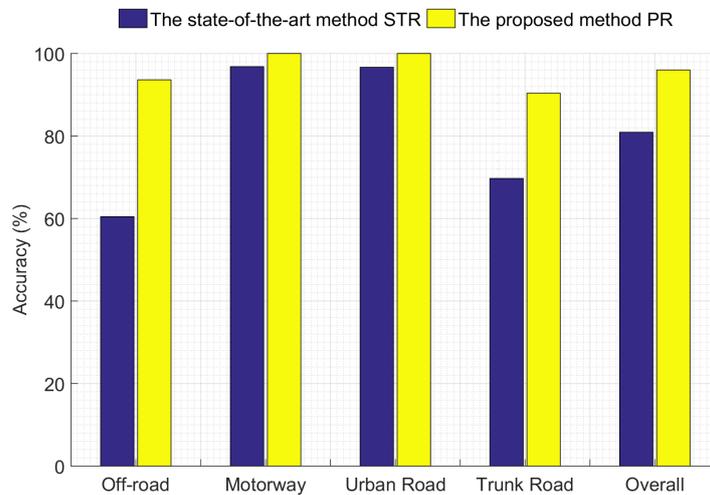


Figure 5.3: Road-type classification results using both methods.

Table 5.2: Classification results in terms of classification accuracy percentage

R. types	Proposed method	[Mioulet et al., 2013]
Off-road	<b>96.8 %</b>	$61.3 \pm 1.75\%$
Motorway	<b>99.7 %</b>	$91.8 \pm 2.15\%$
Urban road	<b>100 %</b>	$92.6 \pm 0.45\%$
Trunk road	<b>92.2 %</b>	$74.4 \pm 2.88\%$
Overall	<b>97.2 %</b>	$80.025 \pm 2.26\%$

Table 5.3: Confusion matrix of the proposed method

Actual \ <i>Predictd</i>	Off-road	Motorway	Urban road	Trunk road
Off-road	0.96875	0	0	0.03125
Motorway	0	0.9975	0.0025	0
Urban road	0	0	1	0
Trunk road	0	0.05125	0.02625	0.9225

Table 5.4: Confusion matrix of [Mioulet et al., 2013]

Actual \ <i>Predictd</i>	Off-road	Motorway	Urban road	Trunk road
Off-road	0.613	0.00075	0.167375	0.218875
Motorway	0.06075	0.917875	0.021375	0
Urban road	0.032875	0.00175	0.9255	0.039875
Trunk road	0.129	0.081125	0.04575	0.744125



Figure 5.4: Road-class labels.

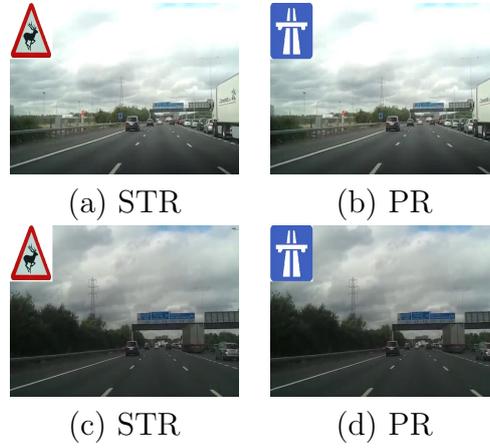


Figure 5.5: Visual comparison of motorway example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.

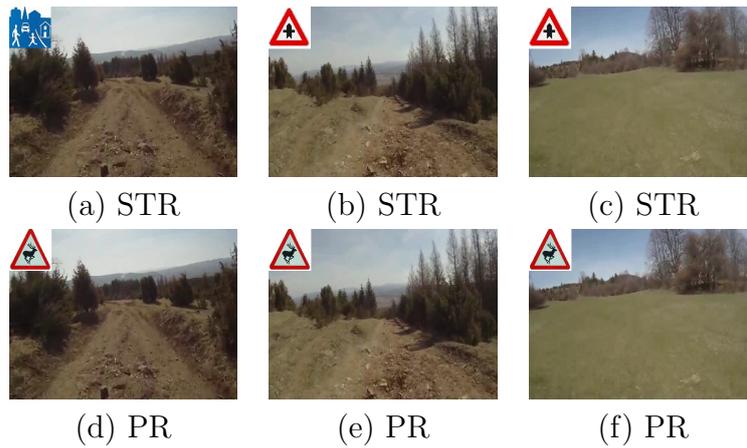


Figure 5.6: Visual comparison of off-road example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.

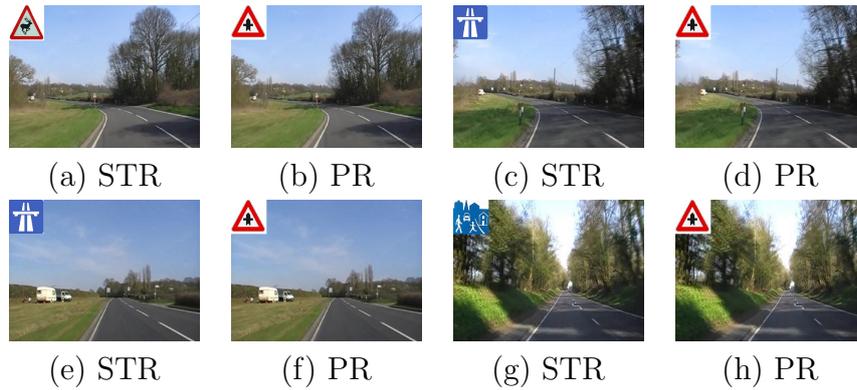


Figure 5.7: Visual comparison of trunk example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.

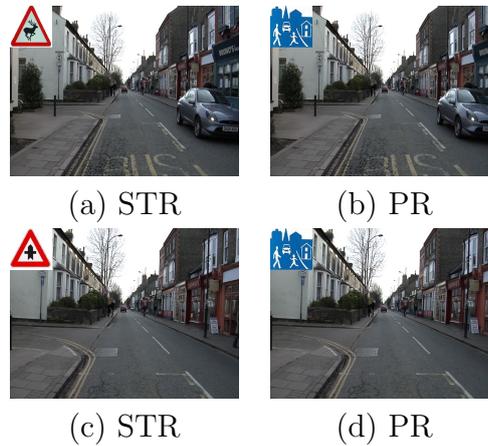


Figure 5.8: Visual comparison of urban example, between the state-of-the-art method STR [Mioulet et al., 2013] and the proposed method PR.

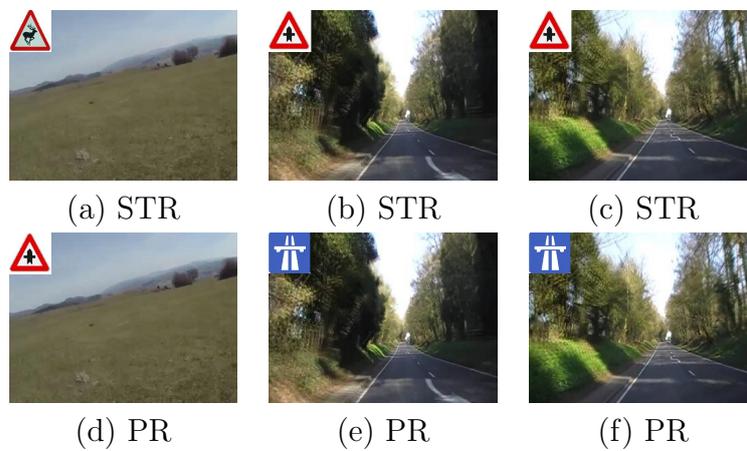


Figure 5.9: Examples of misclassified frames by the proposed method, and correctly classified frames by STR [Mioulet et al., 2013]

## 5.5 Summary

This chapter proposes a new online method for classifying road types, which is based on video segmentation and evolving GMMs. All information from the visual content of the scene is used, without giving any priority to spatial or perceptual areas of the scene. A four-class problem with four different road types are considered.

For testing and comparison with the state-of-the-art method in [Mioulet et al., 2013], several video sequences of different road types are selected, each sequence comprising several hundred frames. This dataset is split into training and testing parts. Finally, both the proposed method and the method in [Mioulet et al., 2013] are implemented on the above dataset.

The experimental results demonstrate that the proposed method outperforms the method in [Mioulet et al., 2013] in terms of both classification accuracy per road type and overall classification accuracy. This achievement can be attributed to using the information from all areas of the frames. However, the method in [Mioulet et al., 2013] is faster than the proposed method, as it provides results in real time, due to exploiting hardware facilities for filtering. The proposed method is online, and there is room for optimisation, such as by using different platforms instead of MATLAB, thereby exploiting the facilities of hardware implementation such as PGA or FPGA. The future work will investigate the optimisation of the proposed method and conduct testing on more datasets. Lastly, this method can be used in the automatic risk-assessment framework proposed in this thesis. As mentioned in Chapter 3, road type is considered a subclass of road environmental risk, and not all road types have the same contribution in creating risk,

because each road type requires a specific driving behaviour. Consequently, classifying road types accurately leads to improvements in the assessment of risk level.

# ONLINE ROAD DETECTION

Road detection is essentially important for both robotic and autonomous vehicles [Wang et al., 2015b]. Road detection helps researchers understand road scenes in terms of safety, and it provides crucial support for road users' safe access to the road. Thus, road detection is a key component of the body of the risk-assessment framework proposed in Chapter 3.

As explained in Chapter 2, the existence of a monocular camera is gradually becoming standard for modern vehicles, with an increasing number of vehicles being equipped with dashboard cameras. Hence, an on-line vision-based road detection method is crucial—a method to handle problems in the detected road region caused by shadows, illuminations and unusual road-shapes. Therefore, in this chapter, a new method for road detection is proposed. This online, model-based method uses video captured by a single video camera. The goal of the method is to acquire the road information from the input frame, which is incorporated into the proposed risk-assessment framework.

The main contribution of this chapter is a novel online, model-based method for road detection that utilises only vision-based features. It first builds an offline threefold generative road model using training data and then uses this model for online road detection. The threefold

road model is based on video segmentation and geometrical cues learnt from prior knowledge of the road’s shape. After the initial detection of the road area, the result is improved using several post-processing steps, such as boundary refinement and region growing. These steps cater for inaccuracies in the detected road region caused by standard research challenges in computer vision, such as shadows, illuminations and unusual road shapes.

Experimental results on the established, publicly available CamVid dataset [Brostow et al., 2008] indicate that the proposed method achieves high accuracy results according to two measures: pixel-wise percentage accuracy and area under the *ROC* curve *AUC*.

This chapter is structured as follows. Section 6.1 provides an overview of the proposed method. The process of building the threefold model, which discriminates between road and nonroad regions, is presented in Section 6.2. The road detection pipeline is described in Section 6.3. Experiments and results are presented in Section 6.4 and the main conclusions of the chapter are summarised in Section 6.5.

## 6.1 Method overview

In this section, a high-level overview of the proposed method for road detection is given. The method consists of two stages. In the first (training) stage, a threefold statistical road model is built using training data, and in the second (detection), the road area in new video frames are detected.

The training stage can be divided into three steps (Figure 6.1). In the first step, the training frames are segmented using the evolving Gaussian mixture model (EvoGMM) algorithm [Kaloskampis and

Hicks, 2014], and models for road and nonroad areas are built from the Gaussians corresponding to each area type. The resulting models for road and nonroad are GMMs. The use of EvoGMM guarantees the compactness of the road and nonroad models, as it merges similar components. This compactness reduces the computational cost of the generation of the initial road region. In the second step, a prior road shape model is built using ground truth road masks. The prior shape model poses geometrical constraints to the detected road region, improving accuracy. Finally, in the third step, a naive outlier filtering mask is built. This mask eliminates false positives located at a long distance from the road region.

The detection stage, which is illustrated in Figure 6.2, can be divided into four steps. In the first step, the initial road region is generated by building a GMM for the new frame and correlating its components to the GMMs representing road and nonroad areas obtained during the training stage using the Bhattacharyya distance [Bhattacharyya, 1943]. In the second step, the boundaries of the initial area are refined using superpixels. In the third step, the prior road shape model is utilised to eliminate inaccuracies within the detected road area caused by illuminations, shadows, etc. Finally, in the fourth step, to handle unusual road shapes, the region growing method is employed.

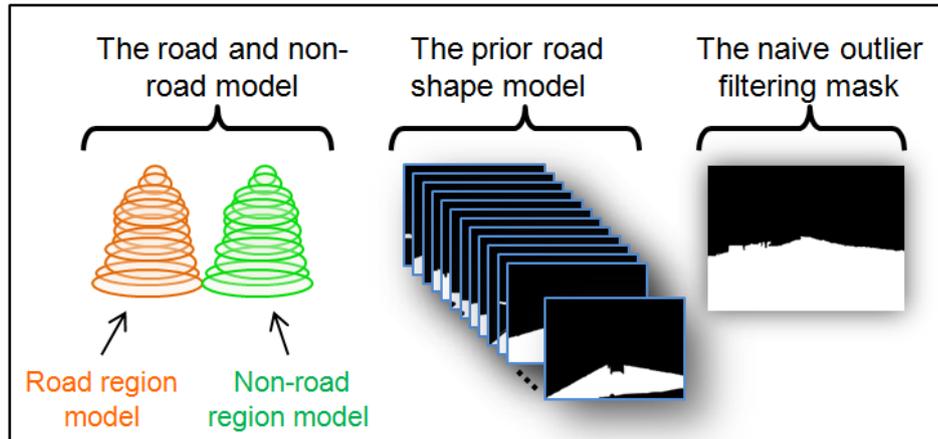


Figure 6.1: The threefold road model.

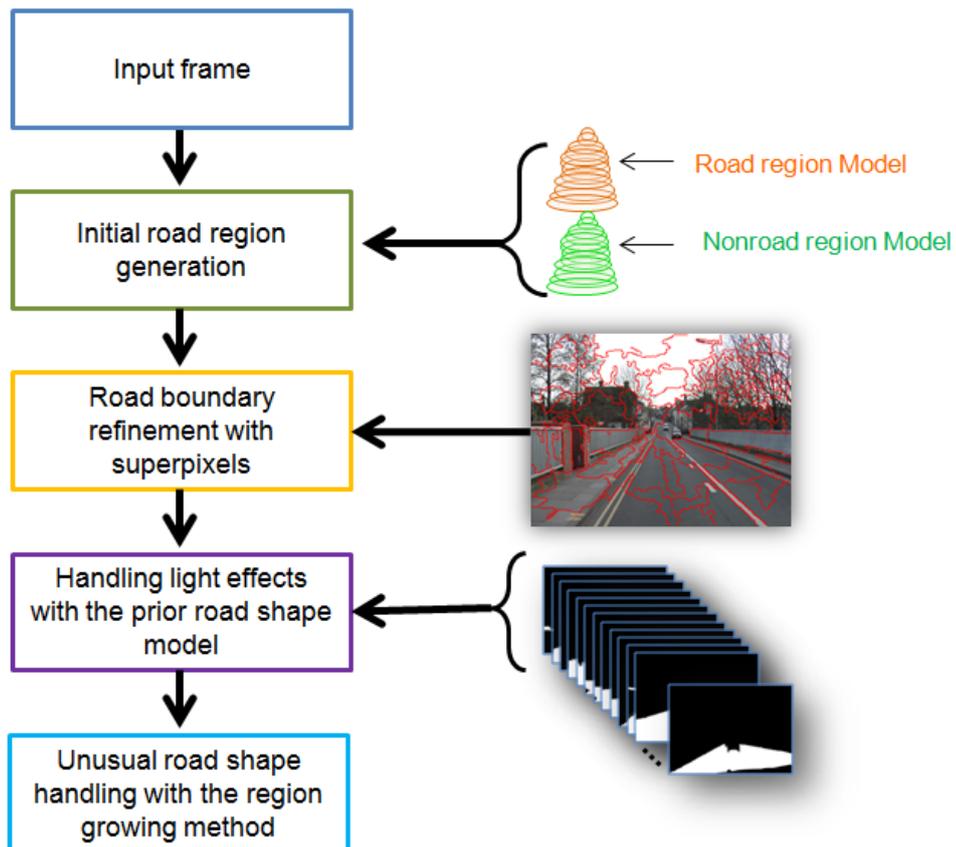


Figure 6.2: Road-detection pipeline.

## 6.2 Building the threefold road model

This section describes the process of building the off-line model of the proposed road detection framework, which will be used in the later on-

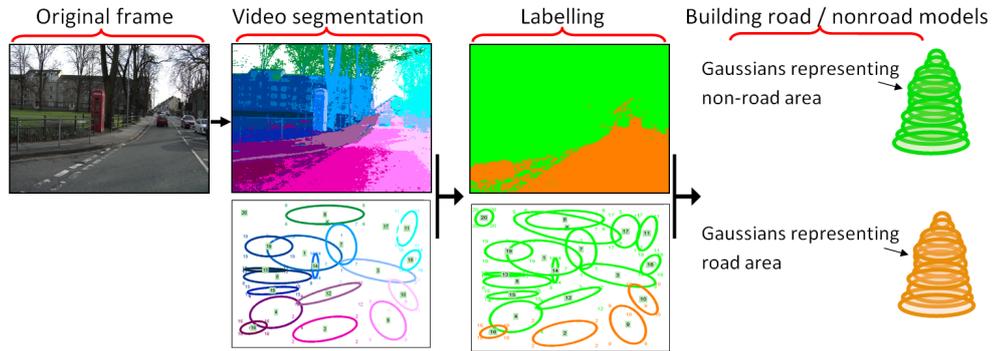


Figure 6.3: Process of building the road and nonroad models with the EvoGMM algorithm.

line road detection phase (Section 6.3). The proposed model is threefold and each of its parts is learned automatically from training data. The model's first part is a statistical model discriminating between road and nonroad regions, the second part is a prior road shape model and the third is a naive outlier filtering mask. Each of these models are discussed in the following subsections.

### 6.2.1 The road and nonroad model

This section describes the process of building the statistical model which discriminates between road and nonroad regions (Figure 6.3). This process is based on the EvoGMM video segmentation algorithm.

The proposed model relies on colour and spatial information from training video frames, which will enable discrimination between road and nonroad regions. As the aim is for the method to work as close as possible to real time, this information should be handled efficiently. For each frame in the training set, features from each of its pixels are extracted. A GMM is built using the features of the frame; after building the GMM, all features extracted from the pixels are discarded. Each frame is, therefore, represented by a GMM rather than its pixel

features, saving a significant amount of computer storage space and memory.

To build a model for the region category road or nonroad, similar to building the road-type models in Chapter 5 Section 5.2, the components of the GMMs, corresponding to this specific region category from all the training frames, are simply concatenated; this means that all GMMs from the training frames are labelled manually as one of two categories, road and nonroad. Finally, the model  $M_i$  for category  $i$  is given by the formula:

$$M_i = \{L_{ik}\}_{k \in \{1, 2, \dots, N_i\}}, \quad (6.1)$$

where  $L_{ik}$  is the  $k^{th}$  Gaussian in  $M_i$ , and  $N_i$  is the total number of Gaussians in  $M_i$ .

### 6.2.2 The prior road shape model

Recent work has shown that the prior road shape provides important geometrical cues, which can be exploited to improve the accuracy of road detection [Álvarez et al., 2014]. Data from geographical information systems (GIS) were used in [Álvarez et al., 2014] to build an online prior road shape model. In this work, an offline method based on vision data is employed to build the prior road shape model. Let  $G$  denote ground-truth masks for the set  $S$  of image sequences with the equation:

$$G = \{g^{(1)}, g^{(2)}, \dots, g^{(m)}\}, \quad (6.2)$$

where  $g^{(n)}, n \in \{1, 2, \dots, m\}$  the ground-truth mask for an image sequence in set  $S$ .

The prior road-shape model  $P$  for the set of image sequences  $S$  is given by the following formula:

$$P = \{G_t^n\}_{t \in \{1, 2, \dots, l\}}^{n \in \{1, 2, \dots, m\}}, \quad (6.3)$$

where  $G_t^n$  is the  $t^{\text{th}}$  ground-truth mask in image sequence  $n$  and  $l$  is the total number of ground-truth masks in the image sequence  $n$ . An example of the prior road-shape model is given in Figure 6.2 (prior road-shape model stage).

### 6.2.3 The naive outlier filtering mask

Apart from providing geometrical cues, the prior road-shape model is also used in this work to build a simple outlier filtering mask. This filtering mask represents the range of all road regions of the training data and is derived from the prior road-shape model's probability map.

The naive outlier filtering mask  $\rho$ , for the prior road shape model  $P$ , is given by:

$$\rho_{xy} = \begin{cases} 1, & \sum_{\gamma=1}^w \{p_{xy}\} > 0 \\ 0, & \sum_{\gamma=1}^w \{p_{xy}\} = 0 \end{cases}, \quad (6.4)$$

where  $\rho_{xy}$  and  $p_{xy}$  the  $xy^{\text{th}}$  pixel value of the naive outlier filtering mask  $\rho$  and the prior road-shape model  $P$ , respectively, and  $w$  is the total number of ground truth masks in the set  $S$  of image sequences and



Figure 6.4: The probability map and global prior mask for CamVid training dataset: (a) the probability map of the prior road shape model; (b) the naive outlier filtering mask.

$\gamma \in \{1, 2, \dots, w\}$ .

The probability map of the of prior-shape model and the naive outlier filtering mask for the CamVid training dataset are shown in Figure 6.4.

### 6.3 Road detection

In this section, the process of detecting the road area is described, using the threefold model of Section 6.2. The pipeline of the method is shown in Figure 6.2. Note that the road detection process is fully online.

#### 6.3.1 Initial road region generation

To detect the road region in an input frame  $f$ , first the frame's model,  $M_f$ , is built using the EvoGMM algorithm and then its components are correlated to the GMMs of the trained model  $M_i$  representing road and nonroad areas obtained during the training stage (Section 6.2.1).

The model  $M_f$  at input frame  $f$  is a GMM, given by:

$$M_f = \{U_{fj}\}_{j \in \{1, 2, \dots, N_f\}}, \quad (6.5)$$

where  $U_{fj}$  is the  $j^{\text{th}}$  Gaussian in  $U_f$ , and  $N_f$  the total number of Gaussians in the model  $M_f$ .

The next step is to estimate the distance between each Gaussian  $U_f$  from the segmented frame  $f$  and the models of road and nonroad regions, using the Bhattacharyya distance [Bhattacharyya, 1943], defined in Section 5.3 Eq. 5.5. Let  $B(U_{fj}, L_{ik})$  be the Bhattacharyya distance between the  $j^{\text{th}}$  Gaussian of the GMM of  $f$  and the  $k^{\text{th}}$  Gaussian of the model  $M_i$ .

The minimum distance  $\beta_{fij}$ , between the  $j^{\text{th}}$  Gaussian in frame  $f$ ,  $U_{fj}$  and the Gaussians in the model  $M_i$  is then calculated:

$$\beta_{fij} = \min \{B(U_{fj}, L_{ik})\}, \quad (6.6)$$

The Gaussians are classified on the basis of the distances. Since the model  $M_i$  includes two categories, road and nonroad (Section 6.2.1), the classification has two possible decision outcomes. The decision is given by the equation:

$$D_{fj} = \arg \min_i \|\beta_{fij}\|, \quad (6.7)$$

where  $D_{fj}$  is the classification outcome for the  $j^{\text{th}}$  Gaussian of frame  $f$  which can be road or nonroad.

Having classified the Gaussians of input frame  $f$ , the initial road region is generated by merging the regions corresponding to the Gaus-

sians classified as road region.

$$I_{xy}^G = \begin{cases} 1, & \{\nu_{xy}\} \in \{R_{fj}\}_{(D_{fj}=r)}, \\ 0, & \{\nu_{xy}\} \in \{R_{fj}\}_{(D_{fj}=n)}, \end{cases} \quad (6.8)$$

where  $I_{xy}^G$  is the  $xy^{th}$  pixel value in the initial road region  $I^G$ ,  $R_{fj}$  is the segmented region in frame  $f$  that corresponds to the  $j^{th}$  Gaussian of frame  $f$ ,  $\nu_{xy}$  is the  $xy^{th}$  pixel in  $R_{fj}$  region and  $r$  and  $n$  are the road- and nonroad region categories, respectively.

The initial road region is filtered using the the naive outlier filtering mask to remove some outliers. The filtering process is given by:

$$I_{xy}^G = \begin{cases} 1, & I_{xy}^G = 1 \text{ and } \rho_{xy} = 1, \\ 0, & I_{xy}^G = 0 \text{ or } \rho_{xy} = 0, \end{cases} \quad (6.9)$$

The process of generating the initial road region is shown in Figure 6.5. The following section shows how the initial result can be improved by employing superpixels.

### 6.3.2 Road-boundary refinement with superpixels

Figure 6.6 illustrates the differences between the initial road region and the ground truth. It can be seen that the boundaries are rough and there are some false positives. To overcome these issues and to improve the result, the boundaries are refined and smoothed by utilising the entropy rate superpixel algorithm [Liu et al., 2011b]. This is achieved by first oversegmenting the video frames to a number of superpixels, and then merging the superpixels that have an overlap rate with the

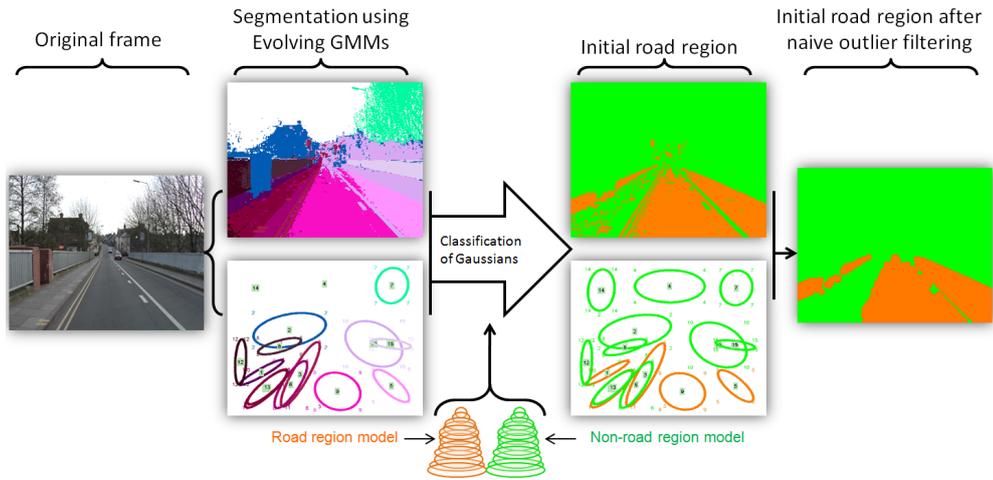


Figure 6.5: Process of generating the initial road region for CamVid testing dataset, frame (*Seq05VD\_03630*).

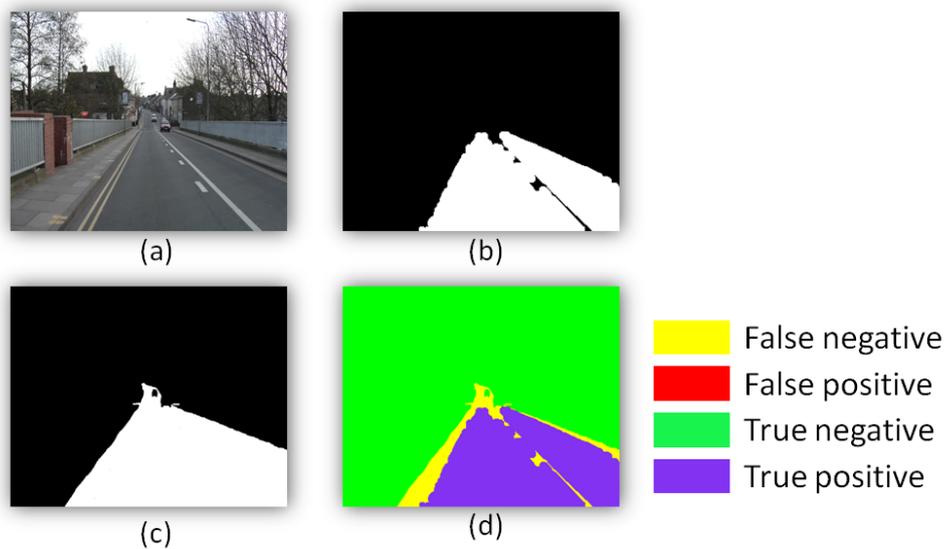


Figure 6.6: Example of the initial road region and its comparison with the ground truth from CamVid testing dataset, frame (*Seq05VD\_03630*): (a) Original frame; (b) The initial road region; (c) Ground truth (d) Initial road region compared to the ground truth.

initial road region. Consequently, the merged area represents the road regions with smooth boundaries.

Superpixel merging is controlled by the region overlapping coefficient  $\alpha$ , which measures the rate of overlap between the superpixels and the initial road region. If  $\theta$  denotes the number of superpixels af-

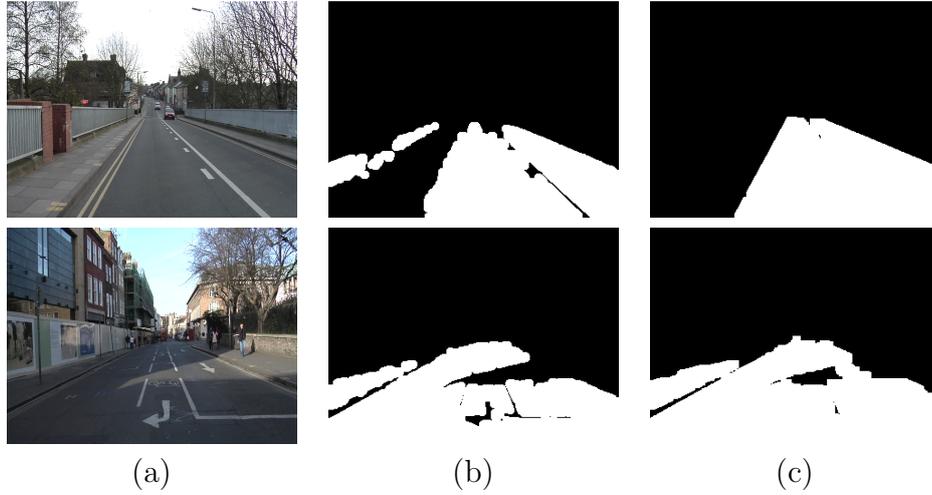


Figure 6.7: Example of road region produced from boundary refinement using superpixels, for CamVid testing dataset, frames (0005VD<sub>0</sub>3630) top row and (0016E5\_002\_08212) bottom row: (a) Original frame; (b) Initial road region; (c) Road region produced from boundary refinement using superpixels.

ter removing nonroad regions, the merging process can be expressed as follows:

$$I_{qxy}^S = \begin{cases} 1 & (\alpha_q \geq \tau)_{q \in \{1, 2, \dots, \theta\}}, \\ 0 & (\alpha_q < \tau)_{q \in \{1, 2, \dots, \theta\}}, \end{cases} \quad (6.10)$$

where  $I_{qxy}^S$  is the  $xy^{th}$  pixel value in region  $q$  of superpixel segmentation road mask  $I^S$  and  $\tau$  is the threshold with  $\tau \in \{0, 0.1, \dots, 1\}$  to control region merging.

Two examples illustrating the output road region after the superpixel step are shown in Figure 6.7.

### 6.3.3 Handling light effects with the prior road shape model

Although the output of the road detection process after the boundary refinement step can be accurate, scene complexity, illuminations, light direction and different levels of shadows might have a negative impact on the result. This is illustrated in Figure 6.7 (bottom row), where

the road region is oversegmented due to shadows. This result can be improved by exploiting the prior road-shape model. The prior road-shape model is, in essence, a concatenation of many different road masks and is built from training data, as explained previously in Section 6.2.2. In this stage, the prior road shape model is used to generate a road-shape for a specific video frame. This is achieved by unifying a number of road shapes within the prior road-shape model that best match the road region generated in the previous step.

To find the best matches, the positive correlation is estimated between each individual road-shape in the model  $P$  (Eq. 6.3) and the road region  $I^S$ . The positive correlation between two shapes can be calculated as follows:

$$\eta(I_1, I_2) = \begin{cases} \Psi(I_1, I_2) & \Psi(I_1, I_2) \geq 0, \\ 0 & \Psi(I_1, I_2) < 0, \end{cases} \quad (6.11)$$

where  $\Psi(I_1, I_2)$  is the Pearson's correlation between  $I_1$  and  $I_2$ , and can be defined as follows:

$$\Psi(I_1, I_2) = \frac{n \sum(I_1 I_2) - (\sum I_1)(\sum I_2)}{\sqrt{(n \sum I_1^2 - (\sum I_1)^2)(n \sum I_2^2 - (\sum I_2)^2)}}, \quad (6.12)$$

where  $n$  is the number of pixels.

The correlation set between the prior-shape model  $P$  and the road region  $I^S$  is given by:

$$\eta_{IP} = \{\eta_{It}\}_{t \in \{1, 2, \dots, w\}}, \quad (6.13)$$

where  $\eta_{It}$  is the positive correlation coefficient between the  $t^{\text{th}}$  road-shape in  $P$  and the road region  $I^S$ , and  $w$  is the total number of road-shapes in  $P$ .

Then, the  $\epsilon$  best-correlated shapes from  $P$  are selected to generate the road shape  $I^P$  by setting a threshold  $\delta$ :

$$I_{xy}^P = \begin{cases} 1, & \sum_{r=1}^{\epsilon} \{\eta_{IP}\} \geq \delta, \\ 0, & \sum_{r=1}^{\epsilon} \{\eta_{IP}\} < \delta, \end{cases} \quad (6.14)$$

where  $I_{xy}^P$  is the  $xy^{\text{th}}$  pixel value of the road shape  $I^P$ .

The richness of the prior road-shape model is the key to control oversegmentation caused by shadows or illuminations. Likewise, the value of  $\delta$  also plays a big role in achieving this. In this experiment, the value of  $\delta$  is set empirically to 15 for the 20 best-correlated shapes ( $\epsilon = 20$ ). Figures 6.9 b and c show the effect of two different  $\delta$ . A higher value of  $\delta$  means involving more prior shapes to build new road shapes. Therefore, it is highly possible to cause overestimation of the road shape.

Two examples of the generated road shape after the step described in this section are shown in Figure 6.8. It can be seen that the accuracy of the road mask has improved compared to the previous step; however, the method is not guaranteed to work as accurately with unusual road shapes. This problem can be tackled with the region growing method,

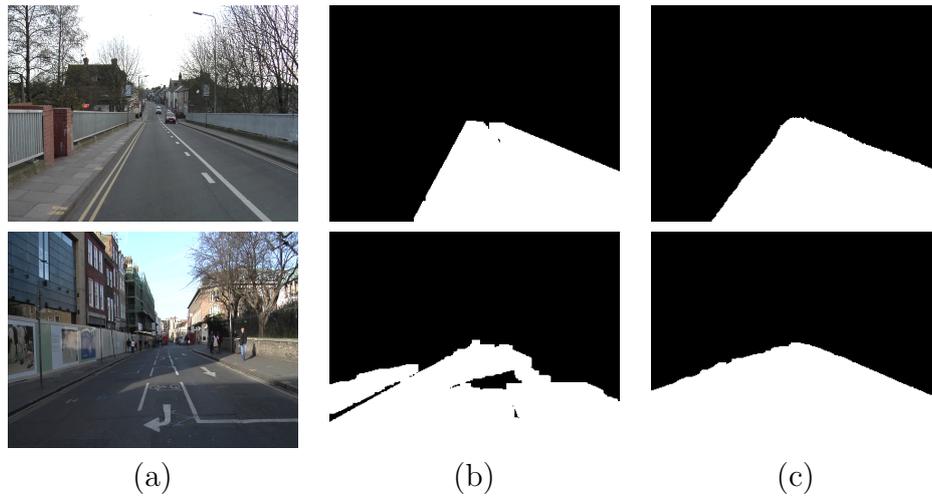


Figure 6.8: Example of road region produced using boundary refinement and the prior road shape model, for CamVid testing dataset, frames (0005VD<sub>0</sub>3630) top row and (0016E5\_002\_08212) bottom row: (a) Original frame; (b) Road region produced from boundary refinement using superpixels; (c) Road region produced using the prior road shape model.

as explained in the next Section.

#### 6.3.4 Unusual road shape handling with the region growing method

This section describes the use of region growing [Adams and Bischof, 1994] within the proposed road detection framework, which is used to deal with unusual road shapes. Region growing is a pixel-based classification method, which utilises the homogeneity between neighbouring pixels to classify them into regions. Although region growing can suffer from shadow and illumination effects [Wen et al., 2008], since both problems can be mitigated at earlier stages of the pipeline, its use at this point is justified. In summary, initial seeds are first selected and the growing process commences with the comparison between the initial seed point and its pixel neighbours on the basis of homogeneity to determine whether they belong to the growing region [Verma et al.,

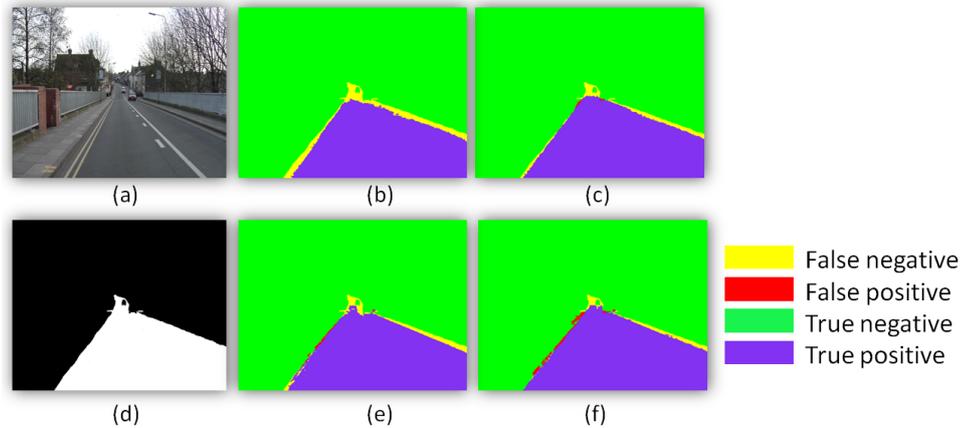


Figure 6.9: Example on the effect of the thresholds and their comparison with the ground truth, for Camvid testing dataset, frame (*Seq05VD\_03630*): (a) Original frame; (b) and (c) Estimated road region using two different values of  $\delta$  (10 and 15, respectively) (d) Ground truth (e) and (f) Estimated road region using two different values of the region growing threshold (1 and 5, respectively).

2011].

In this application, pixels within the road region  $I^P$  estimated in the previous step are selected as initial seeds. All grown regions are unified into a single region, which is the final output of the proposed method. In this experiment, the region-growing threshold is set empirically as 3, Figures 6.9 e and f show the effect of two different thresholds. Higher thresholds lead to more regions being identified as a road, which causes the undersegmentation of the road regions. A sample road region generated after the region-growing step is shown in Figure 6.10.

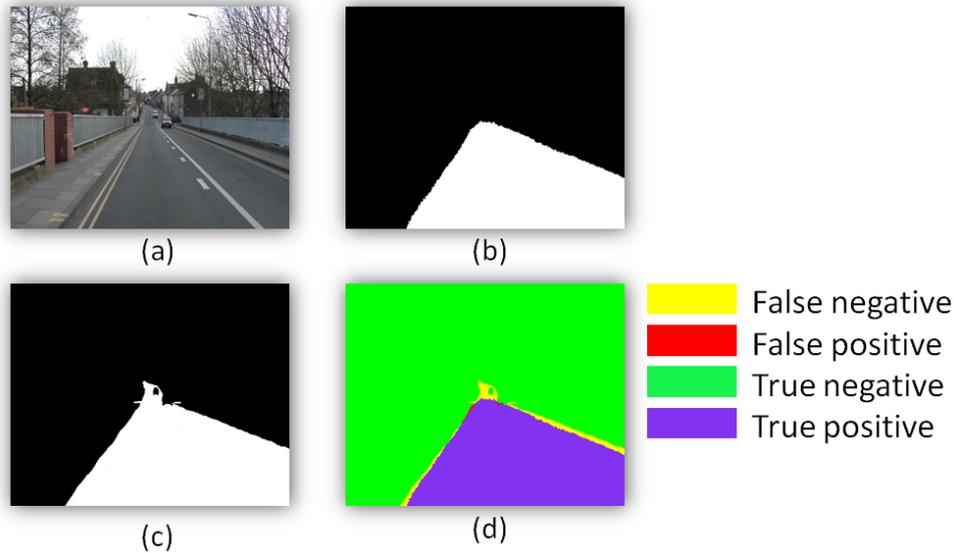


Figure 6.10: Example of estimated road region using region-growing and its comparison with the ground truth, for Camvid testing dataset, frame (*Seq05VD\_03630*):(a) Original frame; (b) The estimated road region using region growing; (c) Ground truth (d) Estimated road region using region growing compared to the ground truth.

## 6.4 Experiments and results

The proposed method is evaluated on the established, publicly available CamVid dataset. The dataset includes daytime and dusk sequences, captured from right-hand drive vehicles and correspond to the driver's perspective. The resolution of the frames is 960X720 pixels. To reduce computational cost, the frames are usually downscaled [Sturgess et al., 2009, Alvarez et al., 2012]. In this experiment, the frames are resized by a scale factor of 1/3. Following the recommendations of the authors of [Brostow et al., 2008], the dataset was split into training and testing subsets. Therefore, as listed in Table 6.1, 468 frames are used to build the offline threefold model of Section 6.2, and 233 frames are used for testing.

Table 6.1: Summary of CamVid dataset

Sequences	Num. Train	Num. Test
EX_1_0001TP	62	62
Ex_2_0006R	71	30
EX_3_0016E_1	90	37
EX_4_0016E_2	125	53
Ex_5_05VD	120	51
All frames	468	233

The CamVid dataset also offers additional motion and 3D structure cues; in this work these features are disregarded, as the proposed method concentrates on visual information. Furthermore, although the CamVid database provides ground truth labels that associate each pixel with one of 32 semantic classes, in an attempt to build a more generative model, this fine-grained annotation is disregarded and only two semantic classes are used, i.e. road and nonroad.

The quantitative evaluation results show that the proposed method achieves high accuracy in road detection, even though it relies solely on visual features. Figure 6.11a shows the comparison between the proposed method and two state-of-the-art methods: the appearance and structure method (AS) [Sturgess et al., 2009] and the segmentation and recognition method (SR) [Brostow et al., 2008] in terms of pixel-wise percentage accuracy. The proposed method outperforms SR and is close to AS. In Figure 6.11 (b) the proposed method is compared against the top-down (TD) and bottom-up (BU) methods from [Al-

varez et al., 2014] and road segmentation (RS) from [Alvarez et al., 2012] using as measure the area under the ROC curve (AUC). State-of-the-art methods achieve higher accuracy, but the proposed method offers competitive results. Note that, with the exception of the BU, TD methods and the proposed algorithm, the remaining methods utilise one or both of the additional cues mentioned earlier (motion and 3D structure and fine-grained annotation). Additionally, the TD method uses an object detector trained on an external dataset. The performance of the proposed method can be improved by including the motion and 3D structure cues, which will be the subject of future work.

Figure 6.12 provides qualitative results of the proposed method for the CamVid dataset. Four examples are given in which road detection is challenging: rows (i) and (ii) feature unusual road shapes; row (iii) illustrates shadow effects; in row (iv) pedestrians occlude the road region and illumination varies near the bottom left corner of the frame. Despite these challenges, the proposed method detects the road with high accuracy.

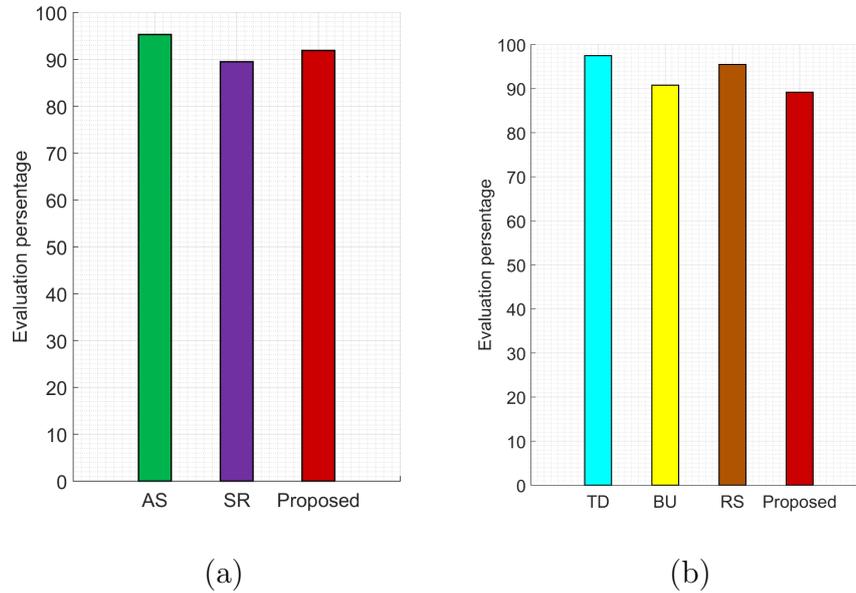


Figure 6.11: Evaluation of road-detection methods and comparison between the state-of-the-art methods and the proposed method according to two measures: (a) Pixel-wise percentage accuracy; (b) Area under the ROC curve (AUC).

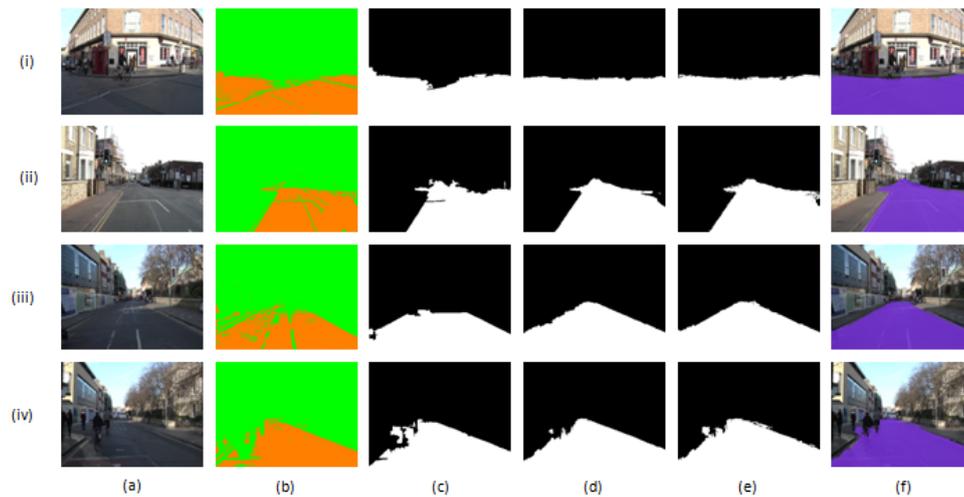


Figure 6.12: Qualitative evaluation of the proposed method for the CamVid dataset: (a) Original; (b) Initial road region; (c) Superpixel boundary refinement; (d) Prior road shape model; (e) Region growing; (f) Superimposition of the output of proposed method on the original frame.

## 6.5 Summary

This chapter proposes a novel method of online road detection that uses as input video captured by a single video camera. The method first

---

builds a threefold statistical road model using training data, and then detects the road area in new frames using this model. The results of initial detection are ameliorated, which handle problems in the detected road region caused by shadows, illuminations and unusual road shapes.

The method is used in the chapter on evaluation of risk-assessment framework (Chapter 7) in order to acquire the road information from the captured frames for the proposed risk-assessment framework.

The performance of the proposed method is evaluated on the CamVid dataset and it is shown that it achieves high accuracy in road detection. The method is online, and there is room for optimisation. Future work will investigate the inclusion of additional cues, such as depth information and GIS in the proposed framework to improve its performance.

# VIDEO-BASED ASSESSMENT OF THE DEGREE OF RISK IN A ROAD SCENE

In this chapter, the pedestrian safety portion of the ontology tool proposed in this thesis (Chapter 3) is evaluated. The main purpose of this evaluation is to explain how the proposed ontology tool can be incorporated into the existing computer vision methods to create a framework that can be used in practice. To do this, different datasets are used to consider two case studies. For this purpose, this chapter proposes a new video dataset that comprises six videos, all of which are taken from YouTube. In addition, five videos from the CamVid dataset [Brostow et al., 2008] are selected, in which all example videos should feature pedestrian behaviour in road scenes with various degrees of risk.

The contributions of this chapter are as follows:

1. A validation framework for the pedestrian safety portion of the proposed ontology tool is developed using two real video datasets.
2. A new real video dataset comprising six videos featuring pedestrian behaviour in road scenes with various degrees of risk is pro-

posed.

This chapter is structured as follows. Semantic feature extraction based on vision methods is investigated in Section 7.1, which presents the computer vision methods used to obtain semantic features, such as road region detection, pedestrian detection and tracking. Moreover, Section 7.2 explains the calculations of the proper semantic attributes, that can be used in this framework, such as speed, location and direction. The combination of this semantic information is demonstrated in Section 7.3. The assessment of pedestrian risks are explained in 7.4. The risk assessment results of the framework evaluation are discussed in Section 7.5, which presents the evaluation results of two datasets. A new dataset is proposed in this chapter that comprises six videos, all of which are taken from YouTube. Five video examples are selected from the CamVid dataset [Brostow et al., 2008]. All example videos feature pedestrian behaviour in road scenes with various degrees of risk. The main conclusions of the chapter are summarised in Section 7.6.

## **7.1 Semantic feature extraction based on computer vision methods**

This framework uses as input video captured by a single monocular video camera. An essential point of using this high-dimensional data is extracting semantic features. Semantic features are sometimes called high-level features, which can be defined as global properties or region-level descriptors related to shape or spatial attitude in a frame [Nixon and Aguado, 2008]. Here, computer vision methods are used to obtain two semantic features: road region in the scene and tracking the

pedestrian in the scene. The measurements for these semantic features, which correspond to key scene entities, are fed to the ontology's tool in the framework, which evaluates the degree of risk in the scene. Each of the features are explained in the following paragraphs.

### 7.1.1 Road region detection

Detecting road region provides valuable semantic information about the road scene, which is one of the important semantic features used in the proposed framework. The main purpose of detecting road region is to check for and determine the location of the pedestrian in the scene, that is, whether the pedestrian is located on the road or not. The extracted information is prepared to be fed to the ontology structure, which infers the degree of risk in the scene.

As explained in Chapter 6, a new online, model-based road detection method is proposed. The method achieves high performance in experiments; therefore, it is used in this framework. An example of the detected road from the CamVid dataset [Brostow et al., 2008] is shown in Figure 7.1.



Figure 7.1: CamVid dataset and two examples of road detection using the proposed method.

### 7.1.2 Pedestrian detection and tracking

The pedestrian is an essential object in the road scene. Therefore, the pedestrian detection and tracking method provides more important semantic information, which is used in this framework. This method uses a cascade object detector, which is a very fast and robust technique [Viola and Jones, 2001b]. The cascade object detector method uses the Viola-Jones algorithm [Viola and Jones, 2001a]. This method can be divided into four stages. The first is feature selection using fast Haar-like block filters. The second is image representation, which is called the integral image stage. The third is a learning algorithm using AdaBoost. The fourth is cascading classifiers [Viola and Jones, 2001b]. This method usually uses many positive and negative images in the training stage. Overall, for both case studies about 1,341 positive and 1,275 negative images are used in the training stage.

This detection method has limitations in determining accurate results, the most common of which are false positive and false negative problems. The first limitation can be overcome by selecting the bounding box based on the confidence at the beginning of the video and then selecting the nearest bounding box for the rest of the video. The second limitation can be overcome by applying an object-tracking method.

Tracking will improve the stability and accuracy of the detection results by predicting a new location for the object when the detection method itself fails to detect the object. In this framework, the Kalman filter [Welch and Bishop, 1995], a filter widely used for tracking, is used. An example of the detected pedestrian from the CamVid dataset [Brostow et al., 2008] is shown in Figure 7.2.



Figure 7.2: CamVid dataset and two examples of pedestrian detection using the Viola-Jones algorithm [Viola and Jones, 2001a].

## 7.2 Speed, location and direction calculation

As explained in Chapter 3, pedestrian speed, location and direction are counted as semantic attributes, which are more important in terms of safety. Therefore, in each frame, once the pedestrians are detected, their location in the scene, speed and direction are estimated. Section 3.2 explains how these features are extracted from frames captured by a monocular camera. In this framework, the same calculation process is applied, and the vertical edges hypothesis is used in determining the pedestrian location.

## 7.3 Data combination

The data obtained from the previous steps are collected and fed to the ontology, which has to be built according to the inputs of the defined inference rules in Section 3.1.3. The combination is made by concatenating all semantic attributes as vectors with size  $1 \times 11$ . Each element of the vector represents a flag of a specific feature. At this stage, the data combination is organised as follows: pedestrian, on the road, on the road edge, on the road side, has no direction, away from the road,

toward the road, away from the road, has high speed, has medium speed, has low speed, has no speed. For example, if the vector value is [1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0], it means there is a pedestrian on the road who is moving away from the road with low speed. When the ontology tool receives the data, on the basis of defined inference rules, it will infer the degree of risk. All results are presented in the next few paragraphs.

#### **7.4 Assessment of pedestrian risk**

In Section 3.1.3, both the risk assessments generated by pedestrian behaviour in the scene and the inference rules are defined. These rules are based on human knowledge and the information from the risk-factor classes. The defined rules are used during the assessment process in the following way: for each frame, the defined rules are applied on the combined data prepared according to the method explained in the previous section. The rules are designed to produce one decision per frame, and these rules are tested using the standard method.

The proposed ontology tool was developed using the *Protégé* resource [pro, 2015], the Pellet reasoner [Dentler et al., 2011] was used to check the consistency of the ontology, and both SPARQL and MATLAB queries were used to query in the testing stage.

#### **7.5 Risk assessment results and evaluation**

In this section, the evaluation of the proposed ontology focuses on the pedestrian safety portion; an evaluation of the complete ontology will be carried out in future works. The experimental results of the proposed

framework are evaluated against the ground truth results.

Furthermore, experimental results are discussed with respect to the ontology's entities that contributed to the reasoning output. For this purpose, the evaluation results of two of the aforementioned datasets are reported.

### 7.5.1 Case study 1: proposed dataset

As explained in the previous sections, the new real video dataset, which comprises six videos, is used as a case study 1. The lengths of the video footages are listed in Table 7.1. For each footage, the frame-based results are compared with the ground truth, and the results of the comparison are depicted in Figures 7.3, 7.4, 7.5 and 7.6. This comparison includes all semantic features used in this framework, as well as the risk assessment. As the figures show, the impact of semantic features when inferring risk assessment varies across time, which means that on the basis of the defined rules, each semantic feature has a key role in some circumstances.

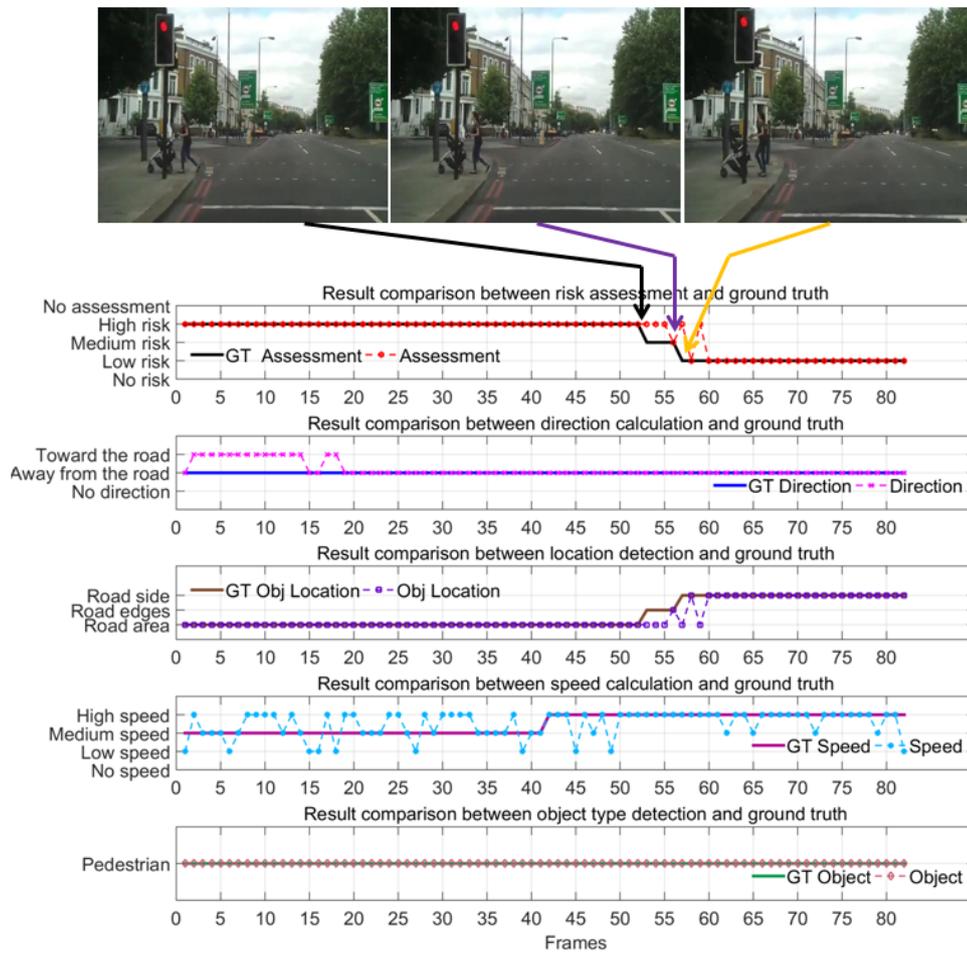


Figure 7.3: Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for case 1 footage 1.

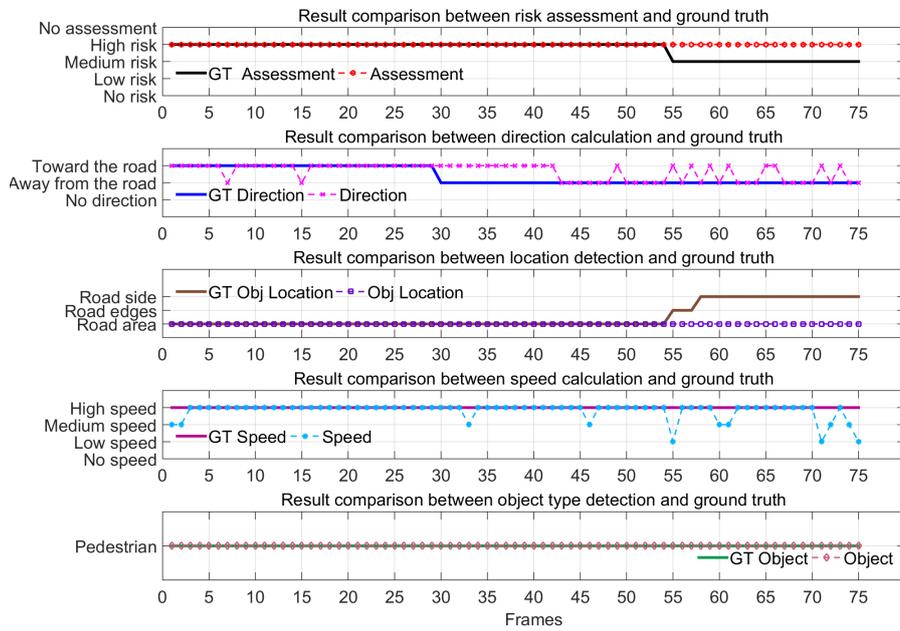


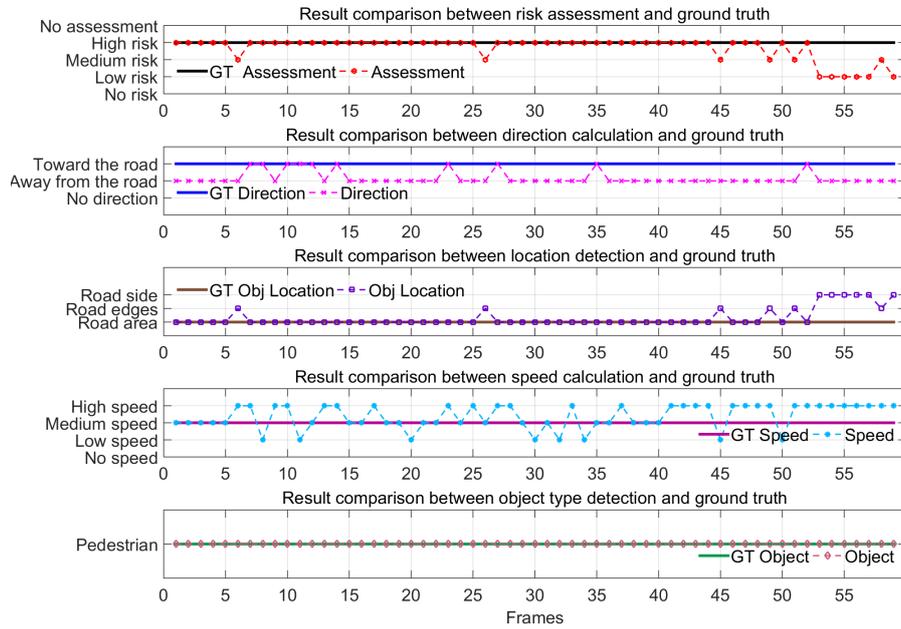
Figure 7.4: Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for case 1 footage 2.

Table 7.1: Summary of proposed dataset for pedestrian behaviour in road scenes

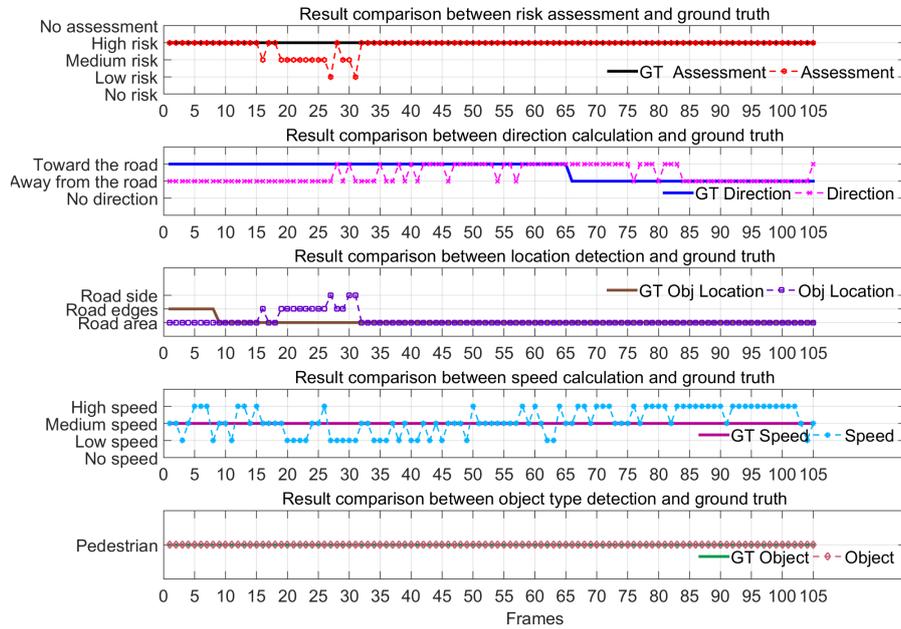
Sequences	Num. frames
case 1 footage 1	83
case 1 footage 2	75
case 1 footage 3	59
case 1 footage 4	105
case 1 footage 5	64
case 1 footage 6	131
All frames	517

For instance, in case 1 footage 1 (Figure 7.3), for frames 56 to 60 the pedestrian location is varied between road area and road edges and becomes a key feature, because this change has a direct impact on the assessment and varies between high risk and low risk. Again, for case 1 footage 3 (Figure 7.5 a), for frames 6, 26, 45, 51, 53 and 58, the pedestrian location is changed from road area to road edges and becomes a key feature; the assessment changes from high risk to medium risk. Likewise, close attention to the comparisons in Figures 7.3, 7.4, 7.5, and 7.6 reveals many key role features can be found.

The evaluation accuracy for all semantic features used in the framework and the risk assessment is shown in Figures 7.7. In conjunction with the previous explanation, the percentage accuracy of semantic features in Figure 7.7 provides clear evidence regarding the key role features. In case 1 footage 1, 2, 3 and 5, the accuracy of the risk

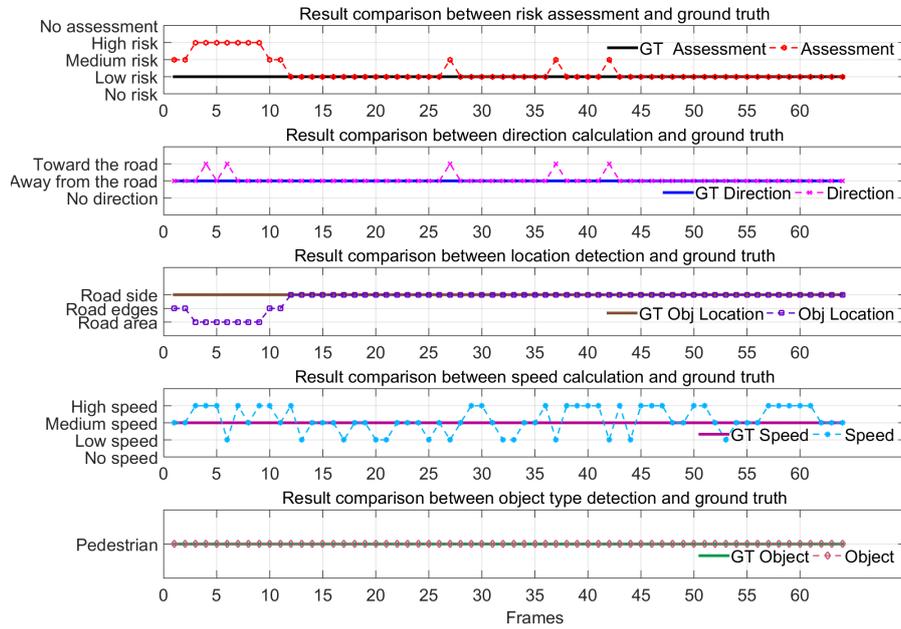


(a) Case 1 footage 3

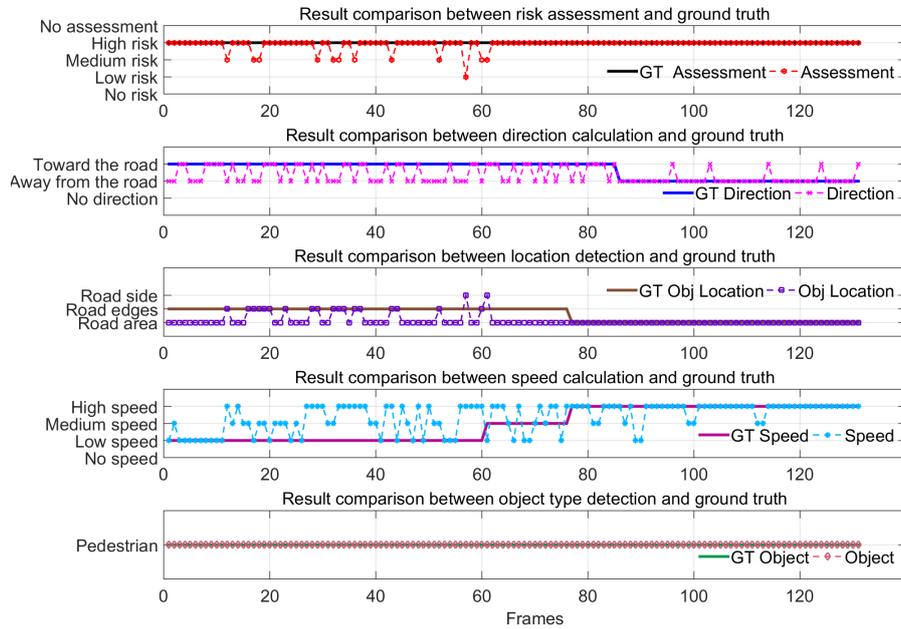


(b) Case 1 footage 4

Figure 7.5: Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for (a) case 1 footage 3 and (b) case 1 footage 4.

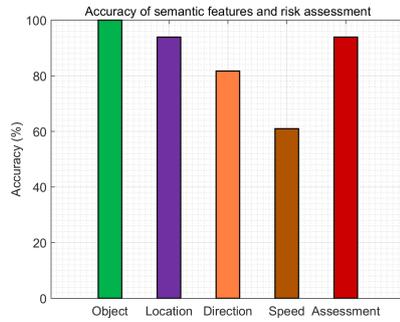


(a) Case 1 footage 5

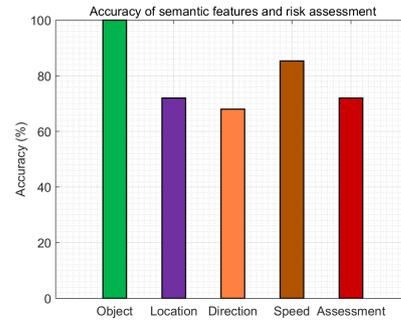


(b) Case 1 footage 6

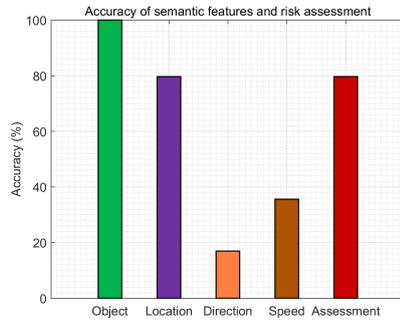
Figure 7.6: Proposed dataset for pedestrian behaviour in road scenes; results of comparison with ground truth for (a) case 1 footage 5 and (b) case 1 footage 6.



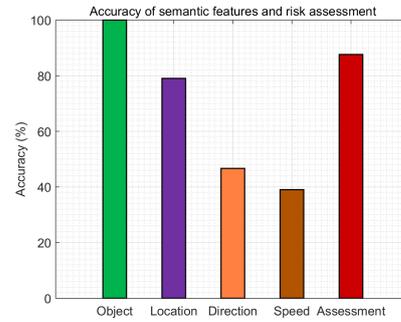
(a) Case 1 footage 1



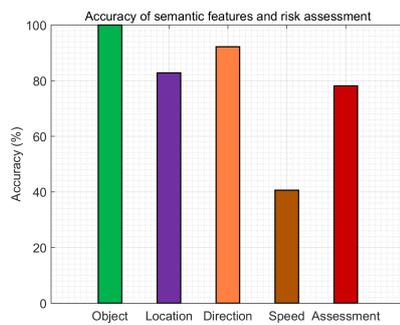
(b) Case 1 footage 2



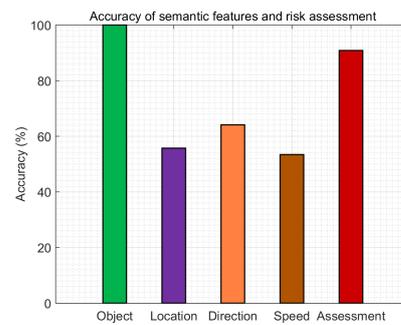
(c) Case 1 footage 3



(d) Case 1 footage 4



(e) Case 1 footage 5



(f) Case 1 footage 6

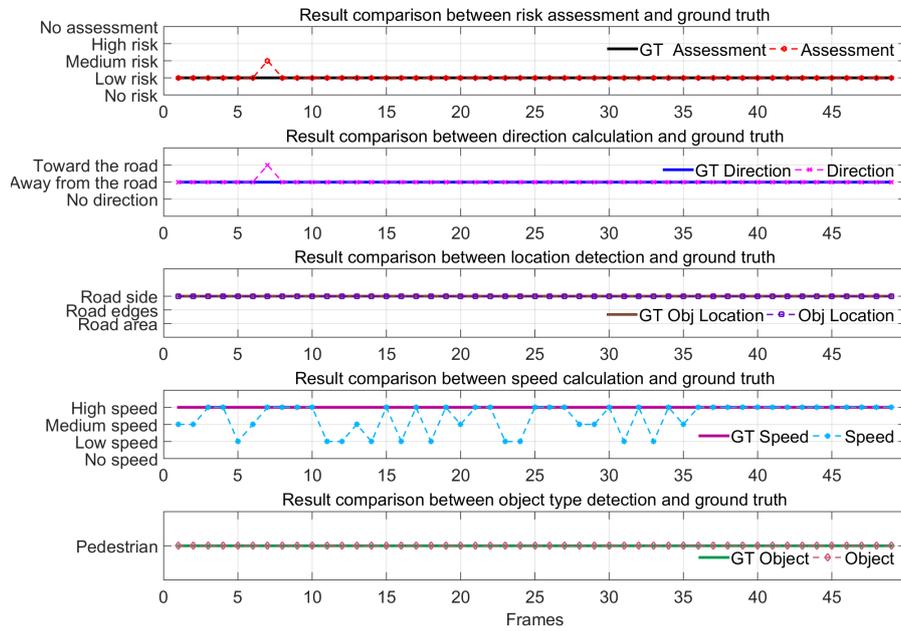
Figure 7.7: Proposed dataset for pedestrian behaviour in road scenes; accuracy results for (a) case 1 footage 1, (b) case 1 footage 2, (c) case 1 footage 3, (d) case 1 footage 4, (e) case 1 footage 5 and (f) case 1 footage 6.

assessment is influenced by the accuracy of location feature than the accuracy of direction and speed (Figure 7.7 a, b, c and e). However, this behaviour changed in case 1 footage 4 and 6 (Figure 7.7 d and f). The accuracy of the risk assessment is much higher than the accuracy of these semantic features—location, direction and speed—which means these three semantic features are accurately obtained, mostly when they have a key role. In this situation, if the key feature is detected accurately, the uncertainty of the other features does not affect the assessment decision.

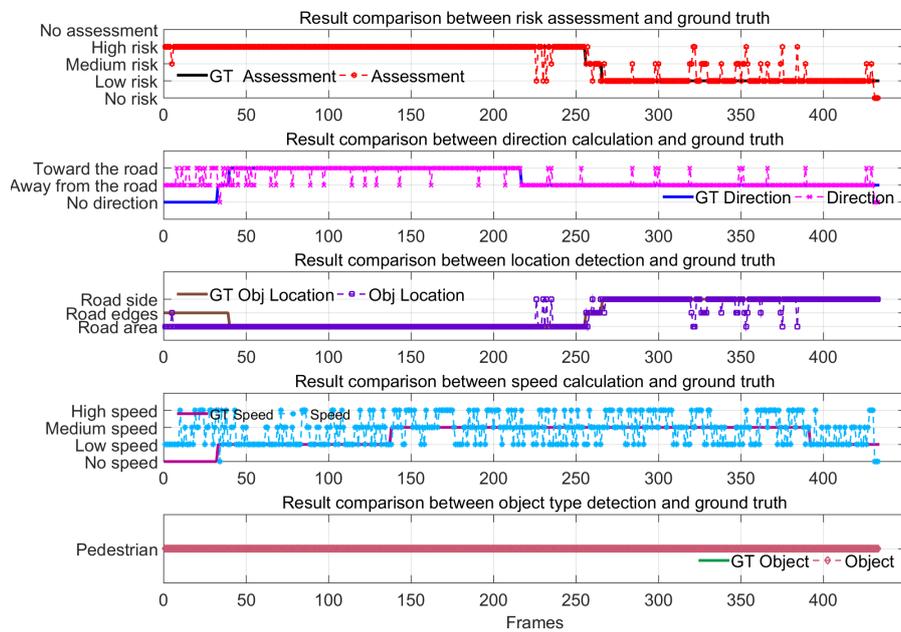
### 7.5.2 Case study 2: CamVid dataset

The proposed framework is evaluated on the CamVid dataset [Brostow et al., 2008], which is a publically available set of videos. Five video examples featuring pedestrian behaviour in road scenes with various degrees of risk are selected. The length of the video examples are listed in Table 7.2. Again, in this case study, for each video footage, the frame-based results are compared with the ground truth and the comparison results are depicted in Figures 7.8, 7.9 and 7.10. Like the first case study, this comparison includes all semantic features that used in this framework, as well as the risk assessment.

Here, the same conclusion of case study 1 can be noticed, for instance, in case 2 footage 1, frame 7 (Figure 7.8 a), the risk assessment is influenced by the direction feature, and in case 2 footage 3 (Figure 7.9 a), the role of the direction feature is more evidenced in frames (3, 7, 12 and 14), but in case 2 footage 4 (Figure 7.9 b), the road feature has the main role to the degree that any changes from direction and speed do not affect the assessment.

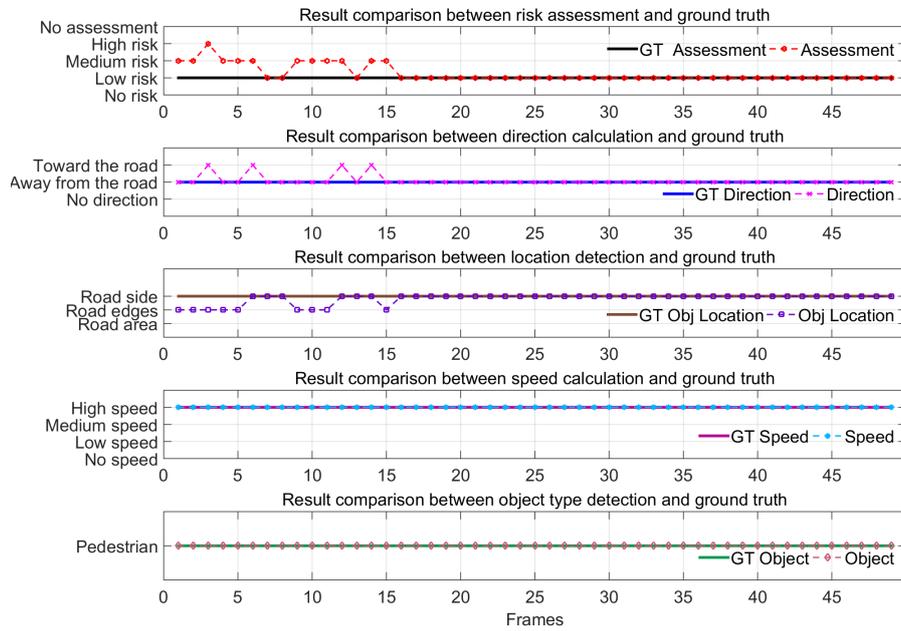


(a) Case 2 footage 1

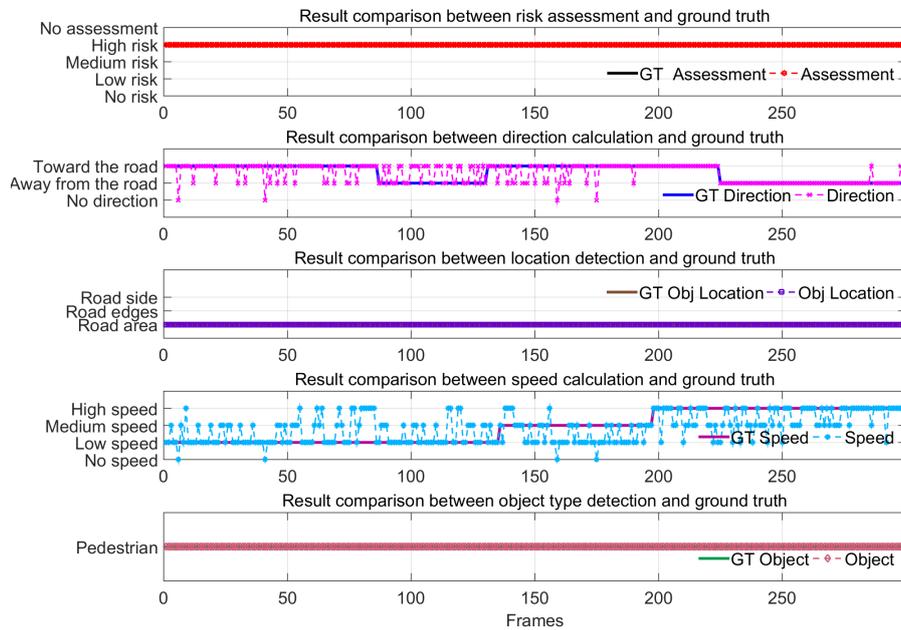


(b) Case 2 footage 2

Figure 7.8: CamVid dataset; results of comparison with ground truth for (a) case 2 footage 1 and (b) case 2 footage 2.



(a) Case 2 footage 3



(b) Case 2 footage 4

Figure 7.9: CamVid dataset; results of comparison with ground truth for (a) case 2 footage 3 and (b) case 2 footage 4.

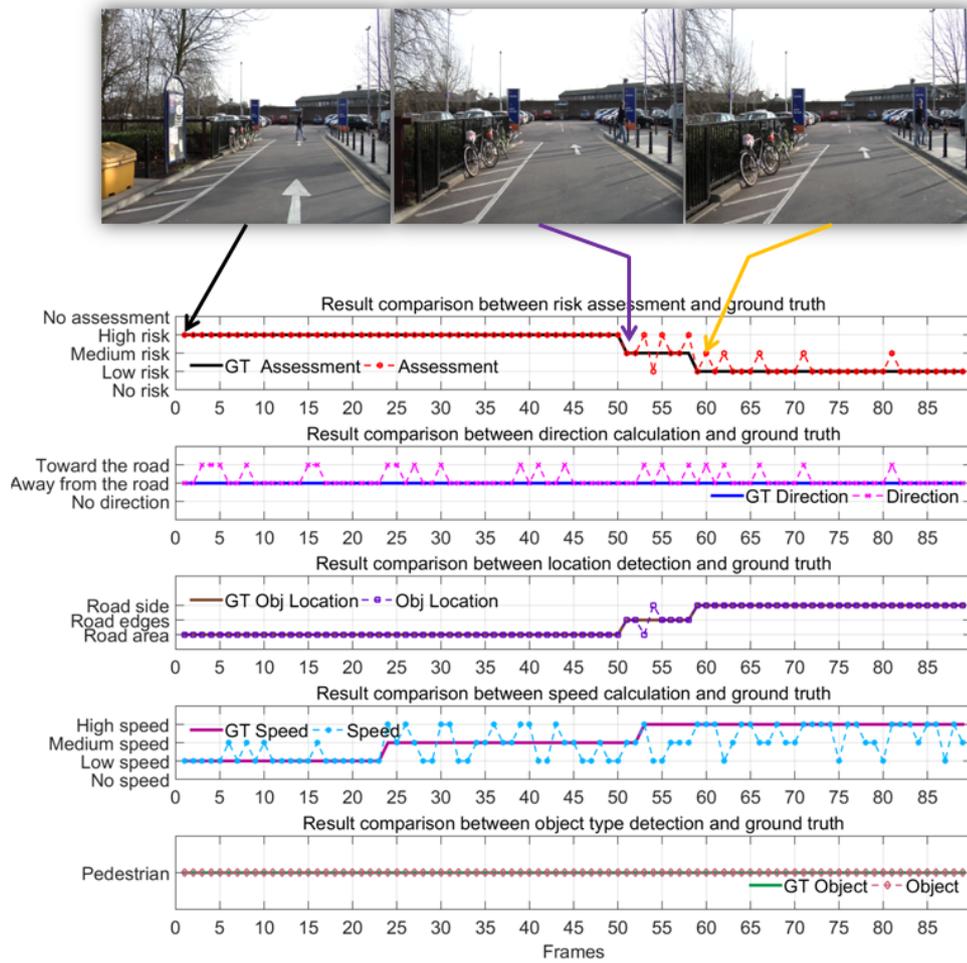


Figure 7.10: CamVid dataset; results of comparison with ground truth for case 2 footage 5.

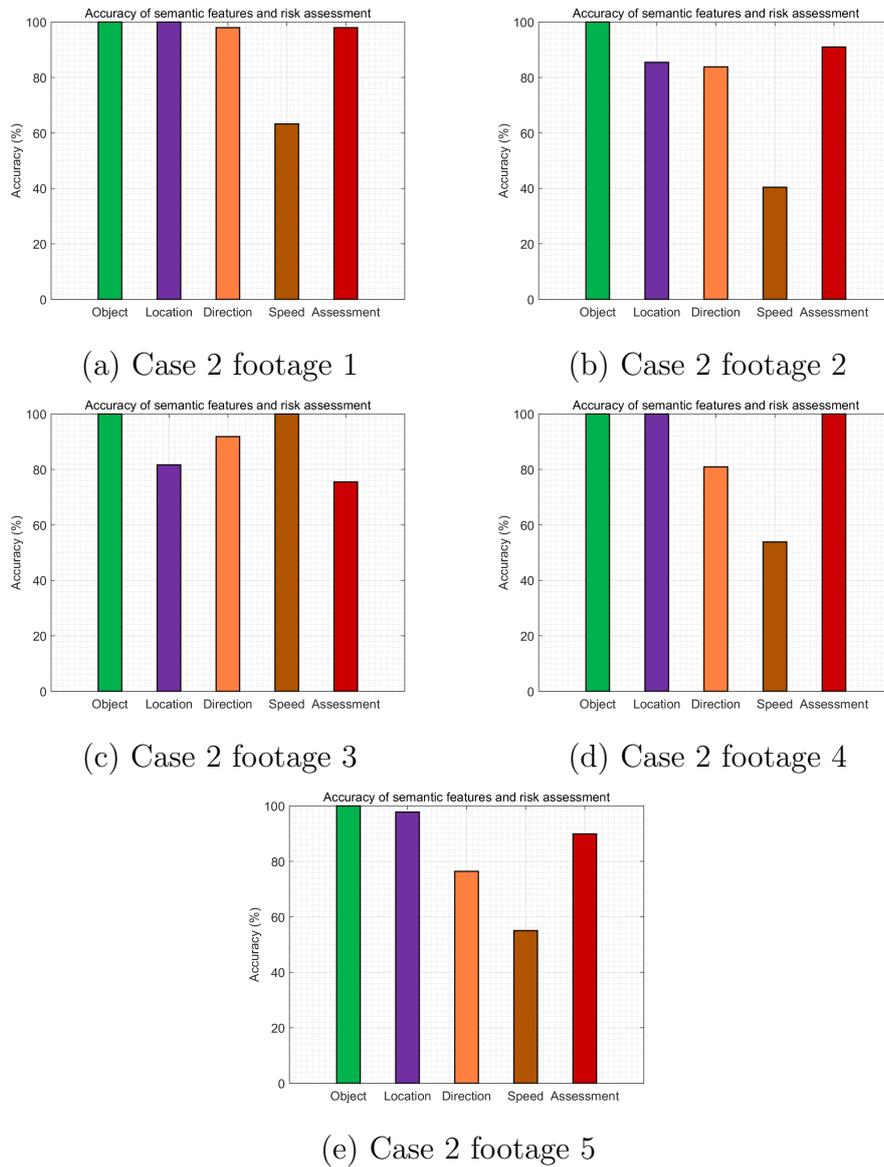


Figure 7.11: Proposed dataset for pedestrian behaviour in road scenes; accuracy results for (a) case 2 footage 1, (b) case 2 footage 2, (c) case 2 footage 3, (d) case 2 footage 4 and (e) case 2 footage 5.

Table 7.2: Summary of CamVid dataset for pedestrian behaviour in road scenes

Sequences	Num. frames
case 2 footage 1	49
case 2 footage 2	433
case 2 footage 3	49
case 2 footage 4	299
case 2 footage 5	89
All frames	919

The evaluation accuracy for all semantic features used in the framework and the risk assessment is shown in Figure 7.11. In conjunction with the previous explanation, the percentage accuracy of semantic features in Figure 7.11 provides clear evidence regarding the key role features. In case 2 footage 1 and 4 (7.11 a and d), the accuracy of the risk assessment is more influenced by the accuracy of the location feature than the accuracy of direction and speed. However, this behaviour changed in case 2 footage 3 (Figure 7.7 c), the accuracy of the risk assessment is lower than the accuracy of these semantic features—location, direction and speed—which means these three semantic features they are not accurately obtained, mostly when they have a key role. In this situation, regardless of whether the rest of the features are correctly obtained, the assessment decision will be made on the basis of the key role features.

### 7.5.3 Overall results

Figure 7.12 shows the accuracy of each case study, the comparison between them and the overall accuracy. In both case studies, the percentage accuracy of detecting semantic features (object, location, di-

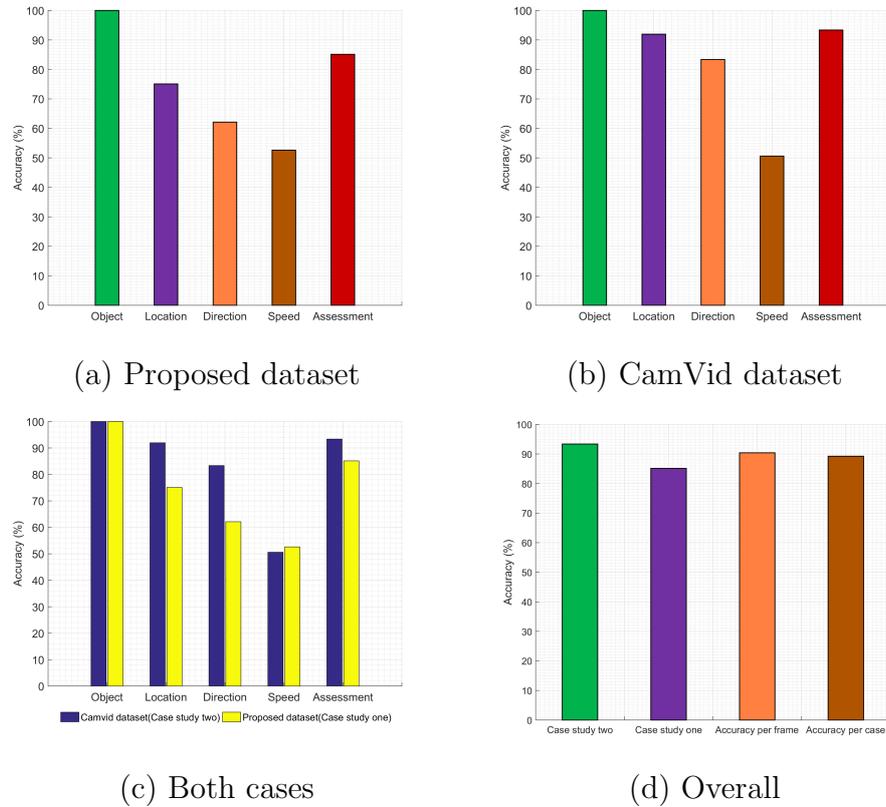


Figure 7.12: Accuracy results for (a) all video footages from the proposed dataset, (b) all video footages from CamVid dataset, (c) both case studies and (d) overall risk assessment for each case study.

rection and speed) differs, but their behaviour is approximately the same. Moreover, the assessment's accuracy overall and in each case study shows that the proposed framework achieves high accuracy. In this way, the assessment's accuracy for the proposed dataset throughout all video footages is 85.1%, but this value is higher for the CamVid dataset (93.3%). The overall accuracy for both frame-based and case-based assessment is high: frame-based accuracy achieves 90.3%, and case-based accuracy achieves 89.2%.

## 7.6 Summary

This chapter proposes a framework for the evaluation of the ontology tool for risk assessment proposed in this thesis. The framework consists of several steps: semantic feature extraction based on computer vision methods, like road region detection and pedestrian detection and tracking; speed, location and direction calculation; data combination; and assessment of pedestrian risk.

The framework's performance is assessed on two datasets: a new dataset proposed in this chapter that comprises six videos, all of which are taken from YouTube, and five video examples selected from the CamVid dataset [Brostow et al., 2008]. Both datasets comprise real-world videos illustrating pedestrian movement.

The experimental results are compared against ground truth, and the percentage accuracy shows that the proposed framework achieves high accuracy in assessing risk resulting from pedestrian behaviour in road scenes.

# CONCLUSIONS AND FUTURE WORK

This chapter concludes the thesis. The main conclusions of this thesis are presented in Section 8.1; Section 8.2 analyses the limitations of the research; and Section 8.3 discusses future work.

### 8.1 Conclusions

Automatic risk assessment and SA are key processes in autonomous driving that support intelligent systems in terms of safety. It is difficult to achieve accurate risk assessment and effective SA in a complex and dynamic scene environment. However, there has been considerable improvement in the field of intelligent transportation infrastructure (ITI) using a variety of sensors, such as GPS, laser sensors, radars and cameras. Due to high cost, complex installation procedures and high computational load, multisensor technology will not become standard for vehicles in the near future. Certain sensors, such as ultrasonic, radar and laser sensors, may additionally suffer from interference problems. In terms of cost and richness of information, it is therefore advisable to use a monocular camera, which is an efficient sensor.

This thesis investigates automatic risk assessment in road scenes using a monocular camera. For this purpose, the methods from two different research areas are exploited and combined into a single framework. To infer semantic information from video data, computer vision methods are used, such as video segmentation, road detection, and pedestrian detection and tracking. Then, on the basis of the semantic information obtained from computer vision methods, the behaviour of the entities in a road scene and the degree of risk of collision in a given scene is inferred automatically. The knowledge-engineering technique of ontology is used for this purpose.

An ontology is designed to represent the various relations between the most important risk factors, risk from object and road environmental risk, which are essential components of the structure of the proposed ontology. The preparation for both components is based on computer vision methods. Moreover, the quality of these methods is important for producing accurate results, especially, video segmentation. There are many different approaches and algorithms for video segmentation; hence, their evaluation is also important for assessing the quality of segmentation results. Nonetheless, little research has focused on the evaluation of video segmentation quality. Therefore, in this thesis a new criteria for high-quality video segmentation to include temporal region consistency is proposed. On the basis of the new criteria, an online method for evaluation of the quality of video segmentation is proposed.

The proposed evaluation method is more consistent than the state-of-the-art method in terms of the perceptual segmentation quality, for both synthetic and real video datasets. For this purpose, a set of syn-

thetic video data is designed and a standard real video dataset is used. Furthermore, using the GMMs video segmentation method, one of the successful video segmentation methods in this area, new methods for both road-type classification and road-detection are proposed. Using the road detection results, along with the pedestrian detection and tracking results, pedestrian speed, location and direction are estimated.

The proposed vision-based road-type classification method achieves higher classification accuracy than the state-of-the-art method for each road type individually, and consequently, it achieves higher overall classification accuracy. This is due to the fact that the state-of-the-art method extracts its features from three predefined subregions in the video frame. However, there is no guarantee that the key information of the scene will always be contained within these regions. Therefore, the proposed method collects features from the entire scene.

The proposed vision-based road-detection method achieves high performance accuracy compared to the state-of-the-art methods, according to two measures: pixel-wise percentage accuracy and area under the *ROC* curve *AUC*. This is due to the use of different methods in the refinement steps, which provides the capacity to address the problems of illumination change, level of shadows and unusual road structures.

A video-based evaluation framework for automatic risk assessment is proposed. At this stage, the framework includes only the pedestrian risk assessment in the road scene. Using the semantic information obtained from computer-vision methods, valuable semantic features such as speed, location and direction are calculated. 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.

The framework's performance is assessed on two datasets: a new dataset proposed in chapter 7 that comprises six videos, all of which are taken from YouTube, and five examples selected from the CamVid dataset. Both datasets comprise real-world videos illustrating pedestrian movement.

The experimental results show that the proposed framework achieves high accuracy in assessing risk resulting from pedestrian behaviour in road scenes.

In conclusion, the achievements of this study reflect several contributions toward the goals and objectives which mentioned in this thesis (Section 1.3). Furthermore, the results of this study confirm the hypothesis of this thesis, that the ontology tool can infer the behaviour of the entities and the degree of risk of collision in a given scene.

## 8.2 Limitations

The limitation of this framework is the accuracy and stability of semantic features of the key entities in the road scene, which is a main factor that can affect the framework's performance, for instance, the accuracy and stability of three semantic attributes—speed, location and direction—have a direct impact on risk-assessment accuracy. However, the achieved accuracy for the automatic risk-assessment framework is promising but insufficient: any false positive or false negative result is crucial. Hence, further improvement is needed to provide more accurate and stable semantic attributes.

### 8.3 Future work

This research has raised many questions in need of further investigation. Further work needs to be done to include all risk factors in the framework, such as vehicle, and its distance, direction and speed.

More accurate and stable information on vulnerable speed, location and direction would help to achieve a higher degree of accuracy in automatic risk assessment. If the accuracy and stability of these semantic attributes are to be improved, a better understanding of road detection, and pedestrian detection and tracking needs to be developed.

It is also recommended that further research be undertaken in the evaluation of video segmentation quality; a number of possible future studies using the proposed criteria of high-quality video segmentation are apparent. It would be interesting to combine the effects of these criteria into one mathematical formula.

---

---

## REFERENCES

- [Cmu, 1997] (1997). CMU dataset. <http://www.cs.cmu.edu/.html/>.
- [pro, 2015] (2015). This work was conducted using the *Protégé* resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health. <http://protege.stanford.edu/>.
- [Adams and Bischof, 1994] Adams, R. and Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647.
- [Allili et al., 2010] Allili, M. S., Ziou, D., Bouguila, N., and Boutemedjet, S. (2010). Image and video segmentation by combining unsupervised generalized gaussian mixture modeling and feature selection. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(10):1373–1377.
- [Alvarez and Lopez, 2011] Alvarez, J. and Lopez, A. (2011). Road detection based on illuminant invariance. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):184–193.
- [Alvarez et al., 2012] Alvarez, J. M., Gevers, T., LeCun, Y., and Lopez, A. M. (2012). Road scene segmentation from a single image. In *Computer Vision–ECCV*, pages 376–389.

- [Alvarez et al., 2009] Alvarez, J. M., Gevers, T., and López, A. M. (2009). Vision-based road detection using road models. In *16th IEEE International Conference on Image Processing (ICIP)*, pages 2073–2076.
- [Álvarez et al., 2014] Álvarez, J. M., López, A. M., Gevers, T., and Lumbreras, F. (2014). Combining priors, appearance, and context for road detection. *IEEE Transactions on Intelligent Transportation Systems*, 15(3):1168–1178.
- [Alvarez et al., 2014] Alvarez, J. M., Salzmann, M., and Barnes, N. (2014). Data-driven road detection. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1134–1141.
- [Arbelaez et al., 2009] Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2009). From contours to regions: An empirical evaluation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2294–2301.
- [Armand et al., 2014] Armand, A., Filliat, D., and Ibanez-Guzman, J. (2014). Ontology-based context awareness for driving assistance systems. In *Intelligent Vehicles Symposium Proceedings, IEEE*, pages 227–233.
- [Askar et al., 2004] Askar, S., Kondratyuk, Y., Elazouzi, K., Kauff, P., and Schreer, O. (2004). Vision-based skin-colour segmentation of moving hands for real-time applications. In *Visual Media Production, 2004.(CVMP). 1st European Conference on*, pages 79–85. IET.
- [Bengler et al., 2014] Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., and Winner, H. (2014). Three decades of driver assis-

tance systems: Review and future perspectives. *Intelligent Transportation Systems Magazine, IEEE*, 6(4):6–22.

[Bertozzi and Broggi, 1998] Bertozzi, M. and Broggi, A. (1998). Gold: a parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, 7(1):62–81.

[Bhattacharyya, 1943] Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109.

[Bonin-Font et al., 2008] Bonin-Font, F., Ortiz, A., and Oliver, G. (2008). Visual navigation for mobile robots: A survey. *Journal of intelligent and robotic systems*, 53(3):263–296.

[Borsotti et al., 1998] Borsotti, M., Campadelli, P., and Schettini, R. (1998). Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19(8):741–747.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[Brendel and Todorovic, 2009] Brendel, W. and Todorovic, S. (2009). Video object segmentation by tracking regions. In *12th IEEE International Conference on Computer Vision*, pages 833–840.

[Broggi and Berte, 1995] Broggi, A. and Berte, S. (1995). Vision-based road detection in automotive systems: A real-time expectation-driven approach. *Journal of Artificial Intelligence Research*, 3(1):325–348.

[Brostow et al., 2008] Brostow, G. J., Shotton, J., Fauqueur, J., and

- Cipolla, R. (2008). Segmentation and recognition using structure from motion point clouds. In *ECCV (1)*, pages 44–57.
- [Brox and Malik, 2010] Brox, T. and Malik, J. (2010). Object segmentation by long term analysis of point trajectories. In *Computer Vision—ECCV*, pages 282–295.
- [Chabrier et al., 2006] Chabrier, S., Emile, B., Rosenberger, C., and Laurent, H. (2006). Unsupervised performance evaluation of image segmentation. *EURASIP Journal on Applied Signal Processing*, 2006:217–217.
- [Chang, 2003] Chang, S.-F. (2003). Content-based video summarization and adaptation for ubiquitous media access. In *Image Analysis and Processing, 2003. Proceedings. 12th International Conference on*, pages 494–496. IEEE.
- [Charron and Hicks, 2010] Charron, C. and Hicks, Y. (2010). An evolving mog for online image sequence segmentation. In *ICIP*, pages 2189–2192.
- [Chen and Corso, 2010] Chen, A. Y. and Corso, J. J. (2010). Propagating multi-class pixel labels throughout video frames. In *Image Processing Workshop (WNYIPW), Western New York*, pages 14–17.
- [Chen and Wang, 2004] Chen, H.-C. and Wang, S.-J. (2004). The use of visible color difference in the quantitative evaluation of color image segmentation. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 3:iii–593.
- [Cheng and Ahuja, 2012] Cheng, H.-T. and Ahuja, N. (2012). Exploiting nonlocal spatiotemporal structure for video segmentation. In *IEEE*

- 
- Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 741–748.
- [Choi and Lee, 2003] Choi, E. and Lee, C. (2003). Feature extraction based on the Bhattacharyya distance. *Pattern Recognition*, 36(8):1703–1709.
- [Choi et al., 2014] Choi, S.-W., Lee, C. H., and Park, I. K. (2014). Scene classification via hypergraph-based semantic attributes subnetworks identification. In *Computer Vision—ECCV*, pages 361–376.
- [Correia and Pereira, 2003] Correia, P. L. and Pereira, F. (2003). Objective evaluation of video segmentation quality. *IEEE Transactions on Image Processing*, 12(2):186–200.
- [Csurka et al., 2004] Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2.
- [DeMenthon and Megret, 2002] DeMenthon, D. and Megret, R. (2002). *Spatio-temporal segmentation of video by hierarchical mean shift analysis*. Computer Vision Laboratory, Center for Automation Research, University of Maryland.
- [Dentler et al., 2011] Dentler, K., Cornet, R., ten Teije, A., and de Keizer, N. (2011). Comparison of reasoners for large ontologies in the OWL 2 EL profile. *Semantic Web*, 2(2):71–87.
- [Dey et al., 2010] Dey, V., Zhang, Y., and Zhong, M. (2010). *A review on image segmentation techniques with remote sensing perspective*.

- [Duda et al., 2012] Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern classification*.
- [Eichner and Breckon, 2008] Eichner, M. and Breckon, T. (2008). Integrated speed limit detection and recognition from real-time video. In *Intelligent Vehicles Symposium*, pages 626–631. IEEE.
- [Endsley, 1995] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64.
- [Erdem et al., 2004] Erdem, C. E., Sankur, B., and Tekalp, A. M. (2004). Performance measures for video object segmentation and tracking. *IEEE Transactions on Image Processing*, 13(7):937–951.
- [Ess et al., 2008] Ess, A., Leibe, B., Schindler, K., and Van Gool, L. (2008). A mobile vision system for robust multi-person tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8.
- [Feng et al., 2012] Feng, M., Jia, P., Wang, X., Liu, H., and Cao, J. (2012). Structural road detection for intelligent vehicle based on a 2d laser radar. In *4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, volume 1, pages 293–296.
- [Fritsch et al., 2013] Fritsch, J., Kuhn, T., and Geiger, A. (2013). A new performance measure and evaluation benchmark for road detection algorithms. In *16th International IEEE Conference on Intelligent Transportation Systems- (ITSC)*, pages 1693–1700.
- [Galasso et al., 2012] Galasso, F., Cipolla, R., and Schiele, B. (2012).

Video segmentation with superpixels. In *Computer Vision–ACCV*, pages 760–774.

[Galasso et al., 2011] Galasso, F., Iwasaki, M., Nobori, K., and Cipolla, R. (2011). Spatio-temporal clustering of probabilistic region trajectories. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1738–1745.

[Galasso et al., 2013] Galasso, F., Nagaraja, N., Cardenas, T., Brox, T., and Schiele, B. (2013). A unified video segmentation benchmark: Annotation, metrics and analysis. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3527–3534.

[Gelasca and Ebrahimi, 2006] Gelasca, E. D. and Ebrahimi, T. (2006). On evaluating metrics for video segmentation algorithms. In *Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, number LTS-CONF-2006-006. INTEL.

[Gökalp and Aksoy, 2007] Gökalp, D. and Aksoy, S. (2007). Scene classification using bag-of-regions representations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8.

[Goldberger and Greenspan, 2006a] Goldberger, J. and Greenspan, H. (2006a). Context-based segmentation of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):463–468.

[Goldberger and Greenspan, 2006b] Goldberger, J. and Greenspan, H. (2006b). Context-based segmentation of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):463–468.

- [Goudail et al., 2004] Goudail, F., Réfrégier, P., and Delyon, G. (2004). Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images. *Journal of the Optical Society of America A (JOSA A)*, 21(7):1231–1240.
- [Gould et al., 2009] Gould, S., Fulton, R., and Koller, D. (2009). Decomposing a scene into geometric and semantically consistent regions. In *IEEE 12th International Conference on Computer Vision (ICCV)*, pages 1–8.
- [Greenspan et al., 2002] Greenspan, H., Goldberger, J., and Mayer, A. (2002). A probabilistic framework for spatio-temporal video representation & indexing. In *Computer Vision (ECCV)*, pages 461–475.
- [Gruber, 1993] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.
- [Grundmann et al., 2010] Grundmann, M., Kwatra, V., Han, M., and Essa, I. (2010). Efficient hierarchical graph-based video segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2141–2148.
- [Guo and Mita, 2009] Guo, C. and Mita, S. (2009). Drivable road region detection based on homography estimation with road appearance and driving state models. In *4th International Conference on Autonomous Robots and Agents (ICARA)*, pages 204–209.
- [Guo et al., 2012] Guo, C., Mita, S., and McAllester, D. (2012). Robust road detection and tracking in challenging scenarios based on Markov random fields with unsupervised learning. *IEEE Transactions on Intelligent Transportation Systems*, 13(3):1338–1354.

- [Han et al., 2012] Han, J., Kim, D., Lee, M., and Sunwoo, M. (2012). Enhanced road boundary and obstacle detection using a downward-looking lidar sensor. *IEEE Transactions on Vehicular Technology*, 61(3):971–985.
- [Haralick and Shapiro, 1985] Haralick, R. M. and Shapiro, L. G. (1985). Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132.
- [He et al., 2013] He, Z., Wu, T., Xiao, Z., and He, H. (2013). Robust road detection from a single image using road shape prior. In *20th IEEE International Conference on Image Processing (ICIP)*, pages 2757–2761.
- [Hearst, 1992] Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pages 539–545. Association for Computational Linguistics.
- [Hoogs and Collins, 2006] Hoogs, A. and Collins, R. (2006). Object boundary detection in images using a semantic ontology. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 111–111. IEEE.
- [Horrocks et al., 2004] Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosz, B., Dean, M., et al. (2004). Swrl: A semantic web rule language combining OWL and RuleML. *W3C Member submission*, 21:79.
- [Hu et al., 2014] Hu, X., Rodriguez, F., and Geppert, A. (2014). A

- multi-modal system for road detection and segmentation. In *IEEE Intelligent Vehicles Symposium Proceedings*, pages 1365–1370.
- [Huang et al., 2009] Huang, Y., Liu, Q., and Metaxas, D. (2009). Video object segmentation by hypergraph cut. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1738–1745.
- [Hulsen et al., 2011] Hulsen, M., Zollner, J., and Weiss, C. (2011). Traffic intersection situation description ontology for advanced driver assistance. In *Intelligent Vehicles Symposium (IV), IEEE*, pages 993–999.
- [Izquierdo and Ghanbari, 2002] Izquierdo, E. and Ghanbari, M. (2002). Key components for an advanced segmentation system. *Multimedia, IEEE Transactions on*, 4(1):97–113.
- [Jansen et al., 2005] Jansen, P., van der Mark, W., van den Heuvel, J., and Groen, F. (2005). Colour based off-road environment and terrain type classification. In *Proceedings IEEE on Intelligent Transportation Systems*, pages 216–221.
- [Jiang et al., 2014] Jiang, Z., Wang, Q., and Yuan, Y. (2014). Adaptive road detection towards multiscale-multilevel probabilistic analysis. In *IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP)*, pages 698–702.
- [Kailath, 1967] Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60.
- [Kaloskampis and Hicks, 2014] Kaloskampis, I. and Hicks, Y. (2014). Estimating adaptive coefficients of evolving GMMs for online video seg-

mentation. In *6th IEEE International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pages 513–516.

[Kannan et al., 2005] Kannan, A., Jojic, N., and Frey, B. (2005). Generative model for layers of appearance and deformation. page 1.

[Kheyrollahi and Breckon, 2012] Kheyrollahi, A. and Breckon, T. P. (2012). Automatic real-time road marking recognition using a feature driven approach. *Machine Vision and Applications*, 23(1):123–133.

[Kumar et al., 2008] Kumar, M. P., Torr, P. H., and Zisserman, A. (2008). Learning layered motion segmentations of video. *International Journal of Computer Vision*, 76(3):301–319.

[Lee and Choi, 2000] Lee, C. and Choi, E. (2000). Bayes error evaluation of the Gaussian ML classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1471–1475.

[Lee et al., 2011] Lee, Y. J., Kim, J., and Grauman, K. (2011). Key-segments for video object segmentation. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1995–2002.

[Leung and Malik, 2001] Leung, T. and Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *International journal of computer vision*, 43(1):29–44.

[Levine and Nazif, 1985] Levine, M. D. and Nazif, A. M. (1985). Dynamic measurement of computer generated image segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):155–164.

- [Levinshtein et al., 2010] Levinshtein, A., Sminchisescu, C., and Dickinson, S. (2010). Spatiotemporal closure. In *Computer Vision–ACCV*, pages 369–382.
- [Lezama et al., 2011] Lezama, J., Alahari, K., Sivic, J., and Laptev, I. (2011). Track to the future: Spatio-temporal video segmentation with long-range motion cues. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [Li and Guo, 2014] Li, X. and Guo, Y. (2014). Multi-level adaptive active learning for scene classification. In *Computer Vision–ECCV*, pages 234–249.
- [Liu et al., 2014] Liu, B., He, X., and Gould, S. (2014). Joint semantic and geometric segmentation of videos with a stage model. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 737–744.
- [Liu et al., 2011a] Liu, C., Yuen, J., and Torralba, A. (2011a). Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):978–994.
- [Liu and Yang, 1994] Liu, J. and Yang, Y.-H. (1994). Multiresolution color image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(7):689–700.
- [Liu et al., 2011b] Liu, M.-Y., Tuzel, O., Ramalingam, S., and Chelappa, R. (2011b). Entropy rate superpixel segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2097–2104.

- [Liu et al., 2013] Liu, W., Zuo, L., Yu, H., Yuan, H., and Zhao, H. (2013). Obstacle detection based on multiple cues fusion from monocular camera. In *16th International IEEE Conference on Intelligent Transportation Systems - (ITSC)*, pages 640–645.
- [Lobato Correia and Pereira, 2004] Lobato Correia, P. and Pereira, F. (2004). Classification of video segmentation application scenarios. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(5):735–741.
- [Ma et al., 2000] Ma, B., Lakshmanan, S., and Hero, A. (2000). Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion. *IEEE Transactions on Intelligent Transportation Systems*, 1(3):135–147.
- [Maedche, 2012] Maedche, A. (2012). *Ontology learning for the semantic web*, volume 665. Springer Science & Business Media.
- [Mak and Barnard, 1996] Mak, B. and Barnard, E. (1996). Phone clustering using the Bhattacharyya distance. In *Proceedings of the Fourth International Conference on Spoken Language, ICSLP 96*, volume 4, pages 2005–2008.
- [Martin et al., 2001] Martin, D., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 416–423.
- [Meilă, 2003] Meilă, M. (2003). Comparing clusterings by the variation of information. In *Learning theory and kernel machines*, pages 173–187.

- [Melo et al., 2006] Melo, J., Naftel, A., Bernardino, A., and Santos-Victor, J. (2006). Detection and classification of highway lanes using vehicle motion trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 7(2):188–200.
- [Mioulet et al., 2013] Mioulet, L., Breckon, T., Mouton, A., Liang, H., and Morie, T. (2013). Gabor features for real-time road environment classification. In *IEEE International Conference on Industrial Technology (ICIT)*, pages 1117–1121.
- [Miranda Neto et al., 2013] Miranda Neto, A., Correa Victorino, A., Fantoni, I., and Ferreira, J. (2013). Real-time estimation of drivable image area based on monocular vision. In *IEEE on Intelligent Vehicles Symposium (IV)*, pages 63–68.
- [Mohammad et al., 2015] Mohammad, M. A., Kaloskampis, I., and Hicks, Y. (2015). Evolving GMMs for road-type classification. In *IEEE International Conference on Industrial Technology (ICIT)*, pages 1670–1673.
- [Morris et al., 1986] Morris, O., Lee, M. d. J., and Constantinides, A. (1986). Graph theory for image analysis: an approach based on the shortest spanning tree. *IEE Proceedings. Part F. Communications, Radar and Signal Processing*, 133(2):146–152.
- [Navigli and Velardi, 2010] Navigli, R. and Velardi, P. (2010). Learning word-class lattices for definition and hypernym extraction. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 1318–1327. Association for Computational Linguistics.

- [Ngan and Li, 2011] Ngan, K. N. and Li, H. (2011). *Video segmentation and its applications*. Springer Science & Business Media.
- [Nixon and Aguado, 2008] Nixon, M. and Aguado, A. (2008). *Feature extraction & image processing*. Academic Press.
- [Ochs and Brox, 2011] Ochs, P. and Brox, T. (2011). Object segmentation in video: a hierarchical variational approach for turning point trajectories into dense regions. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1583–1590.
- [Paris, 2008] Paris, S. (2008). Edge-preserving smoothing and mean-shift segmentation of video streams. In *Computer Vision–ECCV*, pages 460–473.
- [Pelaez et al., 2015] Pelaez, G., Bacara, D., de la Escalera, A., Garcia, F., and Olaverri-Monreal, C. (2015). Road detection with thermal cameras through 3D information. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 255–260.
- [Piccioli et al., 1996] Piccioli, G., Micheli, E. D., Parodi, P., and Campani, M. (1996). Robust method for road sign detection and recognition. *Image and Vision Computing*, 14(3):209–223.
- [Platho et al., 2012] Platho, M., Gross, H., and Eggert, J. (2012). Traffic situation assessment by recognizing interrelated road users. In *15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1339–1344.
- [Pollard et al., 2013] Pollard, E., Morignot, P., and Nashashibi, F. (2013). An ontology-based model to determine the automation level

of an automated vehicle for co-driving. In *16th International Conference on Information Fusion (FUSION)*, pages 596–603.

[Prochazka, 2014] Prochazka, Z. (2014). Improved algorithm for road region segmentation based on sequential monte-carlo estimation. *Computer Science Information Technology*, pages 171–183.

[Ravi et al., 2016] Ravi, D., Bober, M., Farinella, G., Guarnera, M., and Battiato, S. (2016). Semantic segmentation of images exploiting dct based features and random forest. *Pattern Recognition*, 52:260–273.

[Reyes-Aldasoro and Bhalerao, 2006] Reyes-Aldasoro, C. C. and Bhalerao, A. (2006). The Bhattacharyya space for feature selection and its application to texture segmentation. *Pattern Recognition*, 39(5):812–826.

[Rosenberger and Chehdi, 2000] Rosenberger, C. and Chehdi, K. (2000). Genetic fusion: application to multi-components image segmentation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 6, pages 2223–2226.

[Schamm and Zollner, 2011] Schamm, T. and Zollner, J. (2011). A model-based approach to probabilistic situation assessment for driver assistance systems. In *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1404–1409.

[Shi and Malik, 2000] Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

- [Singh and Gupta, 2015] Singh, B. and Gupta, A. (2015). Recent trends in intelligent transportation systems: a review. *Journal of Transport Literature*, 9(2):30–34.
- [Sivic and Zisserman, 2003] Sivic, J. and Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 1470–1477.
- [Spehr et al., 2011] Spehr, J., Rosebrock, D., Mossau, D., Auer, R., Brosig, S., and Wahl, F. (2011). Hierarchical scene understanding for intelligent vehicles. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 1142–1147.
- [Sturgess et al., 2009] Sturgess, P., Alahari, K., Ladicky, L., and Torr, P. H. (2009). Combining appearance and structure from motion features for road scene understanding. In *BMVC*, volume 1, page 6.
- [Taiana et al., 2013] Taiana, M., Nascimento, J., and Bernardino, A. (2013). An improved labelling for the INRIA person data set for pedestrian detection. In Sanches, J., Mic, L., and Cardoso, J., editors, *Pattern Recognition and Image Analysis*, volume 7887 of *Lecture Notes in Computer Science*, pages 286–295.
- [Tang and Breckon, 2011] Tang, I. and Breckon, T. (2011). Automatic road environment classification. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):476–484.
- [Unnikrishnan et al., 2005] Unnikrishnan, R., Pantofaru, C., and Hebert, M. (2005). A measure for objective evaluation of image segmen-

tation algorithms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 34–34.

[Unnikrishnan et al., 2007] Unnikrishnan, R., Pantofaru, C., and Hebert, M. (2007). Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):929–944.

[Vacek et al., 2007] Vacek, S., Gindele, T., Zollner, J., and Dillmann, R. (2007). Situation classification for cognitive automobiles using case-based reasoning. In *Intelligent Vehicles Symposium, IEEE*, pages 704–709.

[Vazquez-Reina et al., 2010] Vazquez-Reina, A., Avidan, S., Pfister, H., and Miller, E. (2010). Multiple hypothesis video segmentation from superpixel flows. In *Computer Vision–ECCV*, pages 268–281.

[Verma et al., 2011] Verma, O., Hanmandlu, M., Susan, S., Kulkarni, M., and Jain, P. (2011). A simple single seeded region growing algorithm for color image segmentation using adaptive thresholding. In *International Conference on Communication Systems and Network Technologies (CSNT)*, pages 500–503.

[Vincenzi et al., 2004] Vincenzi, D. A., Mouloua, M., and Hancock, P. A. (2004). *Human Performance, Situation Awareness and Automation: Current Research and Trends: HPSAA II*, volume 1. Psychology Press.

[Viola and Jones, 2001a] Viola, P. and Jones, M. (2001a). Rapid object detection using a boosted cascade of simple features. In *Proceedings*

---

of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–511.

[Viola and Jones, 2001b] Viola, P. and Jones, M. (2001b). Robust real-time object detection. *International Journal of Computer Vision*, 4:51–52.

[Vogel et al., 2006] Vogel, T., Nguyen, D. Q., and Dittmann, J. (2006). Blobcontours: adapting Blobworld for supervised color and texture-based image segmentation. In *Electronic Imaging*, pages 60730I–60730I.

[Wang and Fremont, 2013] Wang, B. and Fremont, V. (2013). Fast road detection from color images. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 1209–1214.

[Wang et al., 2015a] Wang, D., Hu, K., and Han, X. (2015a). Research on development and model of intelligent transportation systems. In *International Conference on Automation, Mechanical Control and Computational Engineering*.

[Wang et al., 2006] Wang, K., Li, Z., Sun, Y., Qiao, X., and Wang, F.-Y. (2006). An embedded system for vision-based driving environment perception. In *Proceedings of the 2nd IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications*, pages 1–5.

[Wang et al., 2015b] Wang, Q., Fang, J., and Yuan, Y. (2015b). Adaptive road detection via context-aware label transfer. *Neurocomputing*, 158:174–183.

[Welch and Bishop, 1995] Welch, G. and Bishop, G. (1995). An intro-

duction to the kalman filter. Technical report, UNC-CH Computer Science, Chapel Hill, NC, USA, Technical Report 95041.

[Wen et al., 2008] Wen, Q., Yang, Z., Song, Y., and Jia, P. (2008). Road boundary detection in complex urban environment based on low-resolution vision. In *11th Joint International Conference on Information Sciences*, volume 100084, pages 1–7. Atlantis Press.

[Xiao et al., 2010] Xiao, J., Hays, J., Ehinger, K., Oliva, A., and Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3485–3492.

[Xu and Corso, 2012] Xu, C. and Corso, J. J. (2012). Evaluation of super-voxel methods for early video processing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1202–1209.

[Yan et al., 2012] Yan, X., Zhang, H., and Wu, C. (2012). Research and development of intelligent transportation systems. In *11th International Symposium on Distributed Computing and Applications to Business, Engineering Science (DCABES)*, pages 321–327.

[Yang and Zheng, 2015] Yang, M. and Zheng, J. (2015). On-road collision warning based on multiple foe segmentation using a dashboard camera. *IEEE Transactions on Vehicular Technology*, 64(11):4974–4984.

[Yao et al., 2015] Yao, J., Ramalingam, S., Taguchi, Y., Miki, Y., and Urtasun, R. (2015). Estimating drivable collision-free space from monocular video. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 420–427.

- 
- [You et al., 2009] You, C. H., Lee, K. A., and Li, H. (2009). An SVM kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition. *IEEE Signal Processing Letters*, 16(1):49–52.
- [Zhang et al., 2004] Zhang, H., Fritts, J. E., and Goldman, S. A. (2004). An entropy-based objective evaluation method for image segmentation. In *Electronic Imaging*, pages 38–49.
- [Zhang et al., 2008] Zhang, H., Fritts, J. E., and Goldman, S. A. (2008). Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260–280.

## Appendix A

---

# APPENDIX A

The SPARQL query results are shown in this section. The queries are made based on the rule-based cases.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?y ?a
  WHERE {
    ?y rdf:type mam:ReducedVisibility.
    ?res rdf:type mam:RiskAssessment.
    ?res mam:isHighRisk ?a.}

```

	y	a
TheVisibilityIsVeryLow		TheRiskLevelsIsHigh

Execute

Reasoner active  Show Inferences

Figure A.1: SPARQL query, test rule number 1.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?nu ?a
  WHERE {
    ?o rdf:type mam:Vulnerable.
    ?nu rdf:type mam:NonUrbanRoad.
    ?res rdf:type mam:RiskAssessment.
    ?res mam:isHighRisk ?a.}

```

o	nu	a
TheObjectIsVulnerable	TheRoadIsNonUrbanType	TheRiskLevelsIsHigh

Figure A.2: SPARQL query, test rule number 2.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?r ?a
  WHERE {
    ?o rdf:type mam:Vulnerable.
    ?r rdf:type mam:Road.
    ?res rdf:type mam:RiskAssessment.
    ?res mam:isHighRisk ?a.}

```

o	r	a
TheObjectIsVulnerable	OnTheRoad	TheRiskLevelsHigh

Execute

Reasoner active  Show Inferences

Figure A.3: SPARQL query, test rule number 3.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?r ?c ?k ?g ?t ?a
  WHERE {
    ?o rdf:type mam:Vulnerable.
    ?r rdf:type mam:RoadEdge.
    ?t rdf:type mam:AwayFromTheRoad.
    ?g rdf:type mam:BadRoadSurface.
    ?k rdf:type mam:UrbanRoad.
    ?c rdf:type mam:NormalWeather.
    ?res rdf:type mam:RiskAssessment.
    ?res mam:isHighRisk ?a.}

```

o	r	c	k	g	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherisNormal	TheRoadisUrbanType	RoadSurfacelsBad	AwayFromTheRoad	TheRiskLevelsHigh

Figure A.4: SPARQL query, test rule number 4.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?r ?c ?k ?g ?s ?t ?a
  WHERE {
    ?o rdf:type mam:Vulnerable.
    ?r rdf:type mam:RoadEdge.
    ?t rdf:type mam:AwayFromTheRoad.
    ?g rdf:type mam:GoodRoadSurface.
    ?k rdf:type mam:UrbanRoad.
    ?c rdf:type mam:NormalWeather.
    ?res rdf:type mam:RiskAssessment.
    ?o mam:hasHighSpeed ?s.
    ?res mam:isHighRisk ?a.}

```

o	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherisNormal	TheRoadisUrbanType	RoadSurfacelsGood	TheSpeedsHigh	AwayFromTheRoad	TheRiskLevelsHigh

Execute

Reasoner active  Show Inferences

Figure A.5: SPARQL query, test rule number 5.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadEdge.
?k rdf:type mam:AwayFromTheRoad.
?g rdf:type mam:GoodRoadSurface.
?k rdf:type mam:UrbanRoad.
?c rdf:type mam:NormalWeather.
?res rdf:type mam:RiskAssessment.
?O mam:hasMediumSpeed ?s.
?res mam:isHighRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacesGood	TheSpeedIsMedium	AwayFromTheRoad	TheRiskLevelIsHigh

Figure A.6: SPARQL query, test rule number 6.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadEdge.
?k rdf:type mam:AwayFromTheRoad.
?g rdf:type mam:GoodRoadSurface.
?k rdf:type mam:UrbanRoad.
?c rdf:type mam:NormalWeather.
?res rdf:type mam:RiskAssessment.
?O mam:hasLowSpeed ?s.
?res mam:isHighRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacesGood	TheSpeedIsLow	AwayFromTheRoad	TheRiskLevelIsHigh

Execute

Reasoner active  Show Inferences

Figure A.7: SPARQL query, test rule number 7.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadEdge.
?g rdf:type mam:GoodRoadSurface.
?k rdf:type mam:UrbanRoad.
?c rdf:type mam:BadWeather.
?res rdf:type mam:RiskAssessment.
?res mam:isHighRisk ?a.}

```

O	r	c	k	g	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsBad	TheRoadIsUrbanType	RoadSurfacesGood	TheRiskLevelIsHigh

Execute

Reasoner active  Show Inferences

Figure A.8: SPARQL query, test rule number 8.

SPARQL query: [Icons]

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadEdge.
?g rdf:type mam:BadRoadSurface.
?k rdf:type mam:UrbanRoad.
?l rdf:type mam:TowardTheRoad.
?c rdf:type mam:BadWeather.
?O mam:hasNoSpeed ?s.
?res rdf:type mam:RiskAssessment.
?res mam:isHighRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectsVulnerable	OnTheRoadEdge	TheWeatherIsBad	TheRoadIsUrbanType	RoadSurfacetsBad	ThereIsNoSpeed	TowardTheRoad	TheRiskLevelsHigh

Reasoner state out of sync with active ontology  Show Inferences

Figure A.9: SPARQL query, test rule number 9.

SPARQL query: [Icons]

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadEdge.
?g rdf:type mam:GoodRoadSurface.
?k rdf:type mam:UrbanRoad.
?l rdf:type mam:AwayFromTheRoad.
?c rdf:type mam:NormalWeather.
?O mam:hasHighSpeed ?s.
?res rdf:type mam:RiskAssessment.
?res mam:isMediumRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacetsGood	TheSpeedIsHigh	AwayFromTheRoad	TheRiskLevelsMedium

Reasoner state out of sync with active ontology  Show Inferences

Figure A.10: SPARQL query, test rule number 10.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?a
WHERE {
  ?O rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadEdge.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?c rdf:type mam:NormalWeather.
  ?O mam:hasNoSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isMediumRisk ?a.
}

```

O	r	c	k	g	s	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacesGood	ThereIsNoSpeed	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.11: SPARQL query, test rule number 11.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
  ?O rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadEdge.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?t rdf:type mam:AwayFromTheRoad.
  ?c rdf:type mam:NormalWeather.
  ?O mam:hasMediumSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isMediumRisk ?a.
}

```

O	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacesGood	TheSpeedIsMedium	AwayFromTheRoad	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.12: SPARQL query, test rule number 12.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
?O rdfs:type mam:Vulnerable.
?r rdfs:type mam:RoadEdge.
?g rdfs:type mam:GoodRoadSurface.
?k rdfs:type mam:UrbanRoad.
?r rdfs:type mam:AwayFromTheRoad.
?c rdfs:type mam:NormalWeather.
?O mam:hasLowSpeed ?s.
?res rdfs:type mam:RiskAssessment.
?res mam:isMediumRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacesIsGood	TheSpeedIsLow	AwayFromTheRoad	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.13: SPARQL query, test rule number 13.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
?O rdfs:type mam:Vulnerable.
?r rdfs:type mam:RoadSide.
?g rdfs:type mam:GoodRoadSurface.
?k rdfs:type mam:UrbanRoad.
?r rdfs:type mam:AwayFromTheRoad.
?c rdfs:type mam:BadWeather.
?O mam:hasHighSpeed ?s.
?res rdfs:type mam:RiskAssessment.
?res mam:isMediumRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadSide	TheWeatherIsBad	TheRoadIsUrbanType	RoadSurfacesIsGood	TheSpeedIsHigh	AwayFromTheRoad	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.14: SPARQL query, test rule number 14.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?r ?c ?k ?g ?s ?t ?a
WHERE {
  ?o rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadSide.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?t rdf:type mam:AwayFromTheRoad.
  ?c rdf:type mam:BadWeather.
  ?o mam:hasMediumSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isMediumRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectsVulnerable	OnTheRoadSide	TheWeatherIsBad	TheRoadIsUrbanType	RoadSurfacesGood	TheSpeedIsMedium	AwayFromTheRoad	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.15: SPARQL query, test rule number 15.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?r ?c ?k ?g ?s ?t ?a
WHERE {
  ?o rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadSide.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?t rdf:type mam:AwayFromTheRoad.
  ?c rdf:type mam:BadWeather.
  ?o mam:hasLowSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isMediumRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectsVulnerable	OnTheRoadSide	TheWeatherIsBad	TheRoadIsUrbanType	RoadSurfacesGood	TheSpeedIsLow	AwayFromTheRoad	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.16: SPARQL query, test rule number 16.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?g ?s ?t ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadEdge.
?g rdf:type mam:GoodRoadSurface.
?k rdf:type mam:UrbanRoad.
?t rdf:type mam:AwayFromTheRoad.
?c rdf:type mam:NormalWeather.
?O mam:hasLowSpeed ?s.
?res rdf:type mam:RiskAssessment.
?res mam:isMediumRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacelsGood	TheSpeedIsLow	AwayFromTheRoad	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.17: SPARQL query, test rule number 17.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?s ?t ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadEdge.
?g rdf:type mam:GoodRoadSurface.
?k rdf:type mam:UrbanRoad.
?t rdf:type mam:AwayFromTheRoad.
?c rdf:type mam:NormalWeather.
?O mam:hasNoSpeed ?s.
?res rdf:type mam:RiskAssessment.
?res mam:isMediumRisk ?a.}

```

O	r	c	k	g	s	t	a
TheObjectIsVulnerable	OnTheRoadEdge	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacelsGood	ThereIsNoSpeed	AwayFromTheRoad	TheRiskLevelsMedium

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.18: SPARQL query, test rule number 18.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?g ?s ?a
WHERE {
  ?O rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadSide.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?c rdf:type mam:BadWeather.
  ?O mam:hasNoSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isLowRisk ?a.}

```

O	r	c	k	g	s	a
TheObjectsVulnerable	OnTheRoadSide	TheWeatherIsBad	TheRoadIsUrbanType	RoadSurfacesGood	ThereIsNoSpeed	TheRiskLevelsLow

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.19: SPARQL query, test rule number 19.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?t ?s ?a
WHERE {
  ?O rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadSide.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?t rdf:type mam:AwayFromTheRoad.
  ?c rdf:type mam:NormalWeather.
  ?O mam:hasHighSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isLowRisk ?a.}

```

O	r	c	k	g	t	s	a
TheObjectsVulnerable	OnTheRoadSide	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacesGood	AwayFromTheRoad	TheSpeedIsHigh	TheRiskLevelsLow

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.20: SPARQL query, test rule number 20.

```

SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?r ?c ?k ?g ?t ?s ?a
WHERE {
  ?o rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadSide.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?t rdf:type mam:AwayFromTheRoad.
  ?c rdf:type mam:NormalWeather.
  ?o mam:hasMediumSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isLowRisk ?a.
}

```

O	r	c	k	g	t	s	a
TheObjectsVulnerable	OnTheRoadSide	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacetsGood	AwayFromTheRoad	TheSpeedIsHigh	TheRiskLevelsLow

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.21: SPARQL query, test rule number 21.

```

SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?o ?r ?c ?k ?g ?t ?s ?a
WHERE {
  ?o rdf:type mam:Vulnerable.
  ?r rdf:type mam:RoadSide.
  ?g rdf:type mam:GoodRoadSurface.
  ?k rdf:type mam:UrbanRoad.
  ?t rdf:type mam:AwayFromTheRoad.
  ?c rdf:type mam:NormalWeather.
  ?o mam:hasLowSpeed ?s.
  ?res rdf:type mam:RiskAssessment.
  ?res mam:isLowRisk ?a.
}

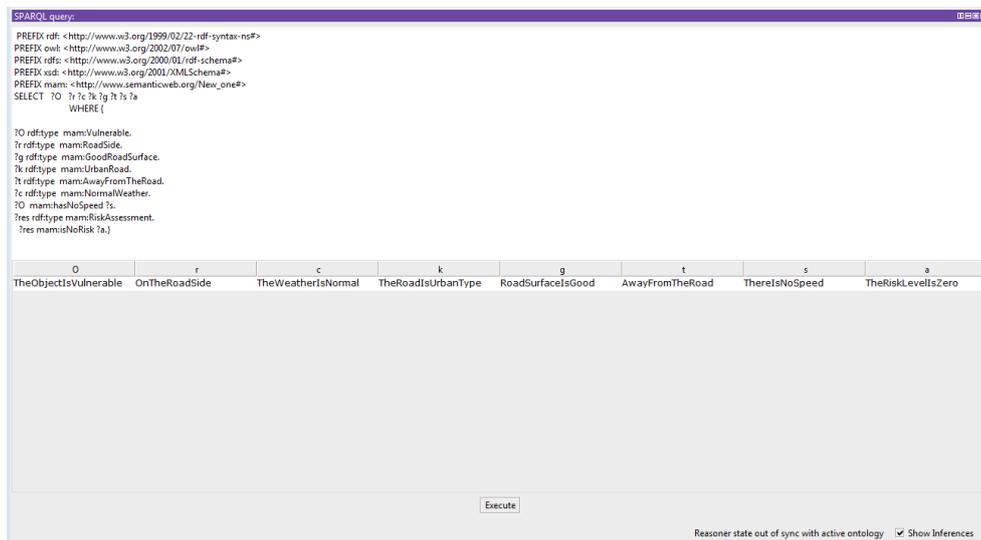
```

O	r	c	k	g	t	s	a
TheObjectsVulnerable	OnTheRoadSide	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacetsGood	AwayFromTheRoad	TheSpeedIsLow	TheRiskLevelsLow

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.22: SPARQL query, test rule number 22.



SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mam: <http://www.semanticweb.org/New_one#>
SELECT ?O ?r ?c ?k ?g ?t ?s ?a
WHERE {
?O rdf:type mam:Vulnerable.
?r rdf:type mam:RoadSide.
?g rdf:type mam:GoodRoadSurface.
?k rdf:type mam:UrbanRoad.
?t rdf:type mam:AwayFromTheRoad.
?s rdf:type mam:NormalWeather.
?O mam:hasToSpeed ?s.
?res rdf:type mam:RiskAssessment.
?res mam:isNoRisk ?a.}
```

O	r	c	k	g	t	s	a
TheObjectIsVulnerable	OnTheRoadSide	TheWeatherIsNormal	TheRoadIsUrbanType	RoadSurfacesGood	AwayFromTheRoad	ThereIsNoSpeed	TheRiskLevelsZero

Execute

Reasoner state out of sync with active ontology  Show Inferences

Figure A.23: SPARQL query, test rule number 23.

# APPENDIX B

The MATLAB query results are shown in this section. Similar to the SPARQL queries, these queries are made based on the rule-based cases. For this purpose, a MATLAB code was developed. First, a menu is displayed (Figure B.1), for the user to choose the query entities, and the button of the selected entity is highlighted (Figure B.2). Then, upon pressing the stop button, a decision menu is displayed, which includes the query entities and the risk assessment.

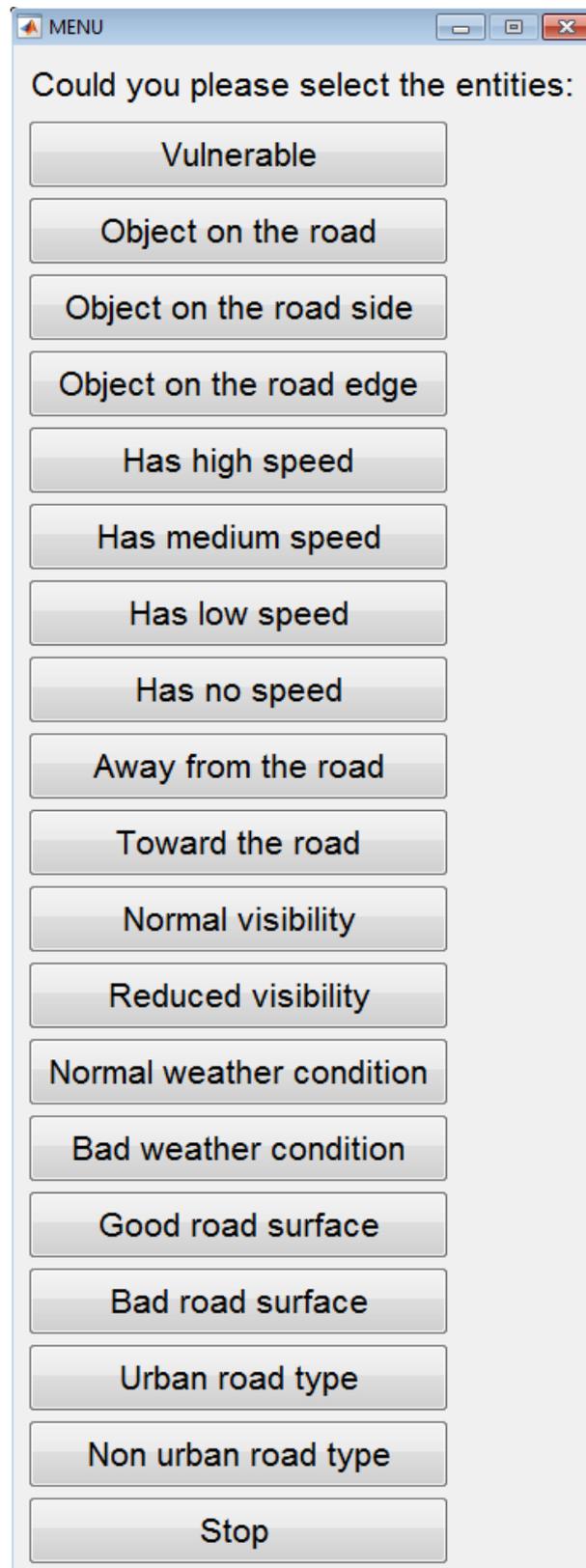


Figure B.1: MATLAB query, main menu to choose the entities.

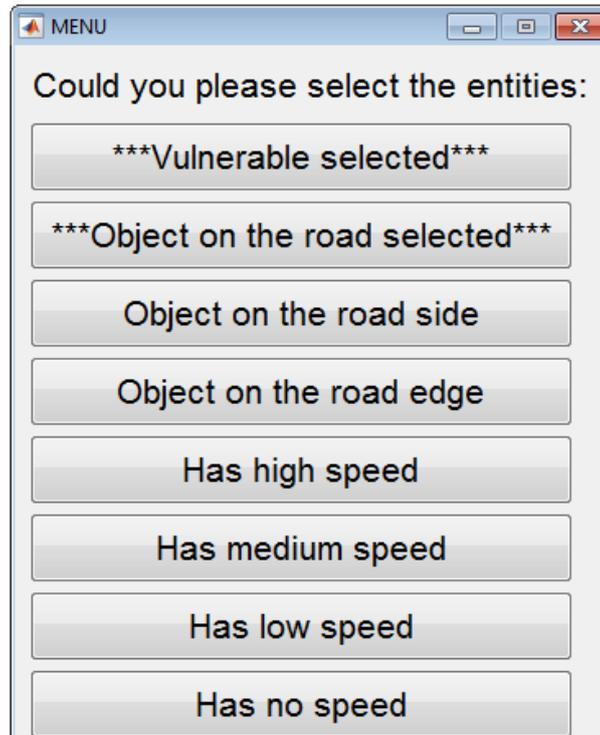


Figure B.2: MATLAB query indicating that the Vulnerable and Object on the road buttons are selected.

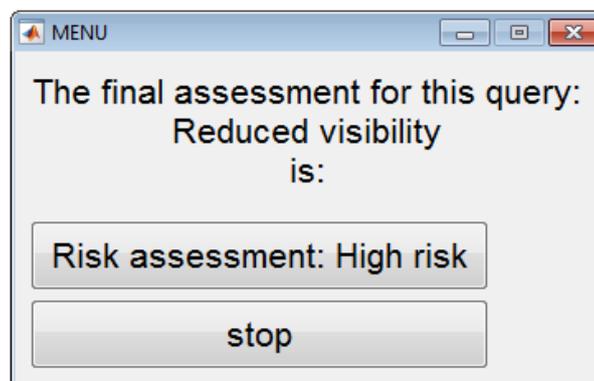


Figure B.3: MATLAB query, test rule number 1.

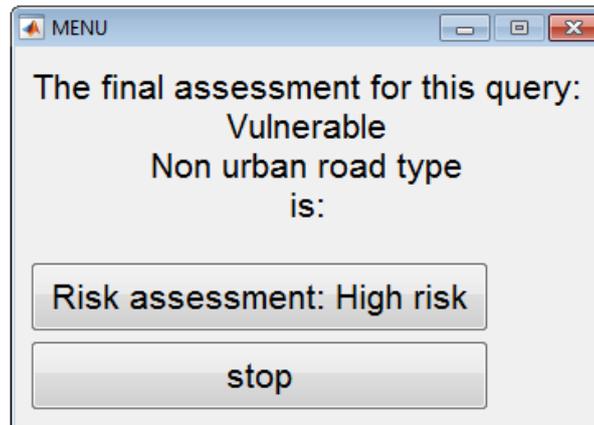


Figure B.4: MATLAB query, test rule number 2.

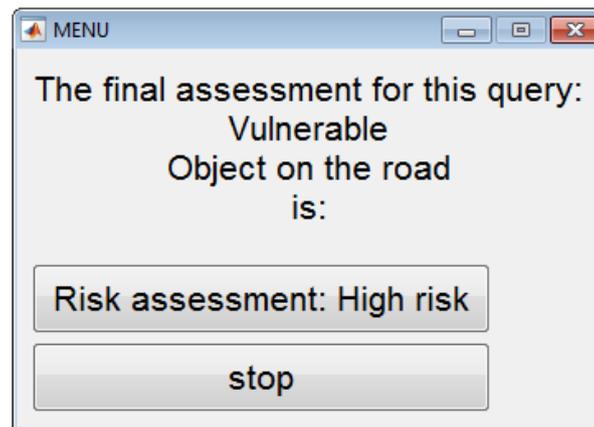


Figure B.5: MATLAB query, test rule number 3.

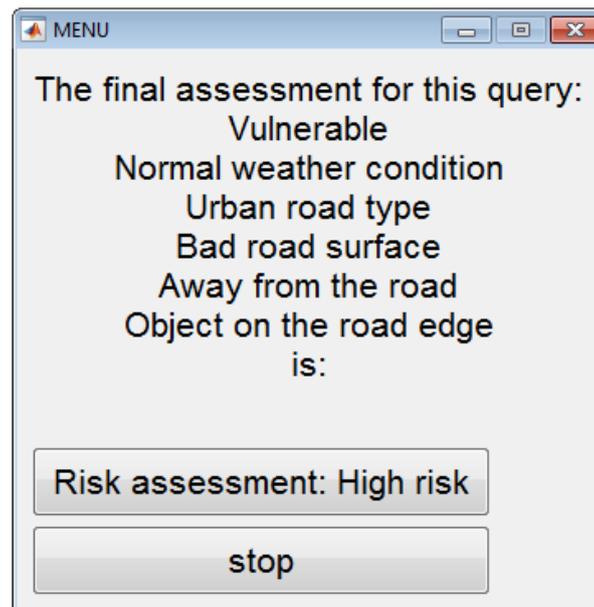


Figure B.6: MATLAB query, test rule number 4.

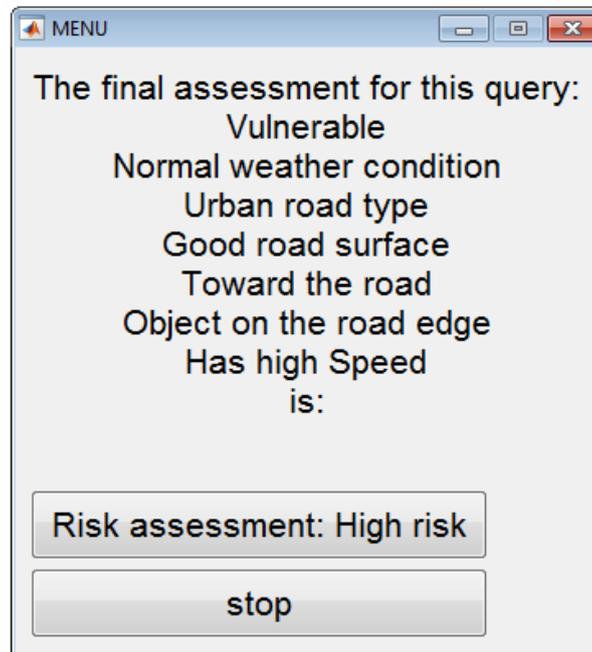


Figure B.7: MATLAB query, test rule number 5.

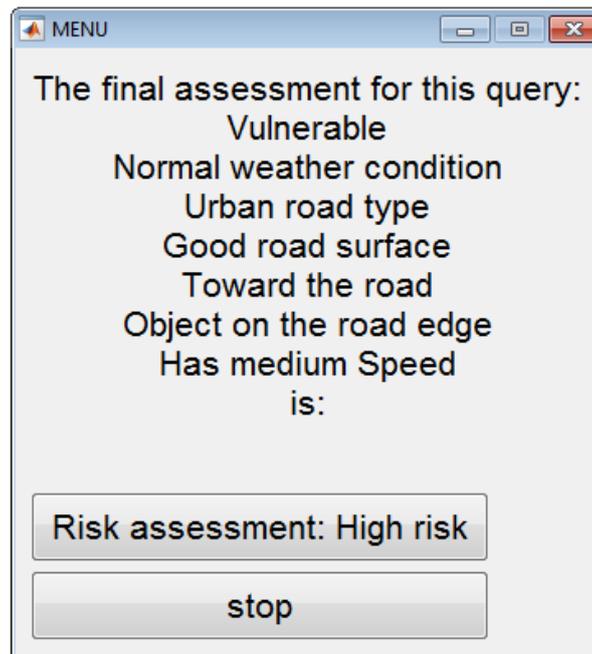


Figure B.8: MATLAB query, test rule number 6.

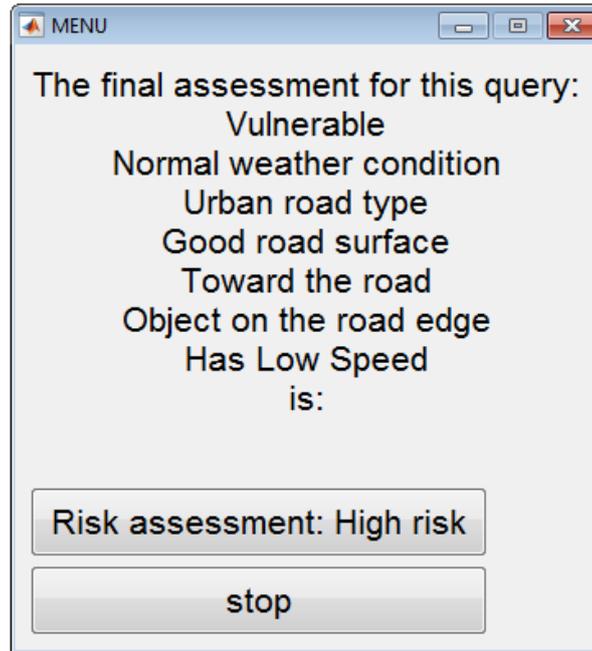


Figure B.9: MATLAB query, test rule number 7.

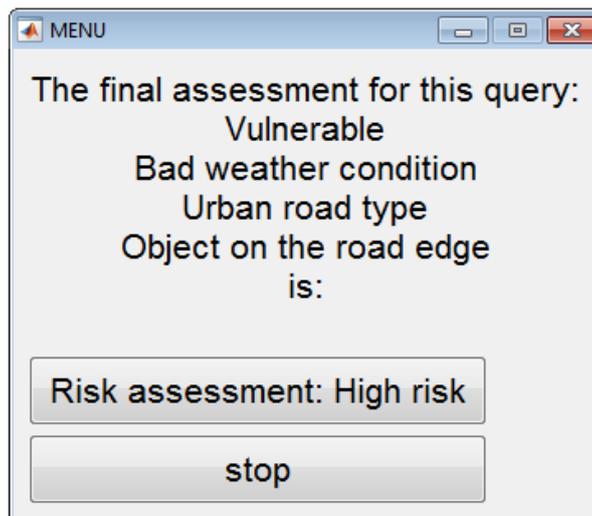


Figure B.10: MATLAB query, test rule number 8.

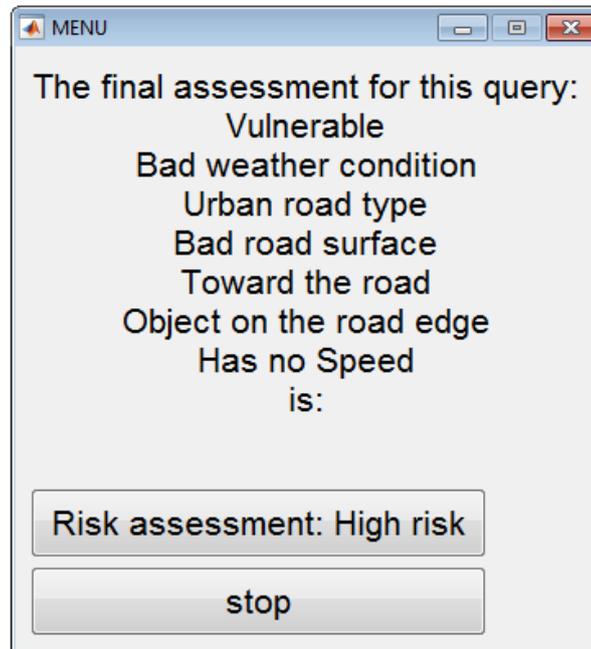


Figure B.11: MATLAB query, test rule number 9.

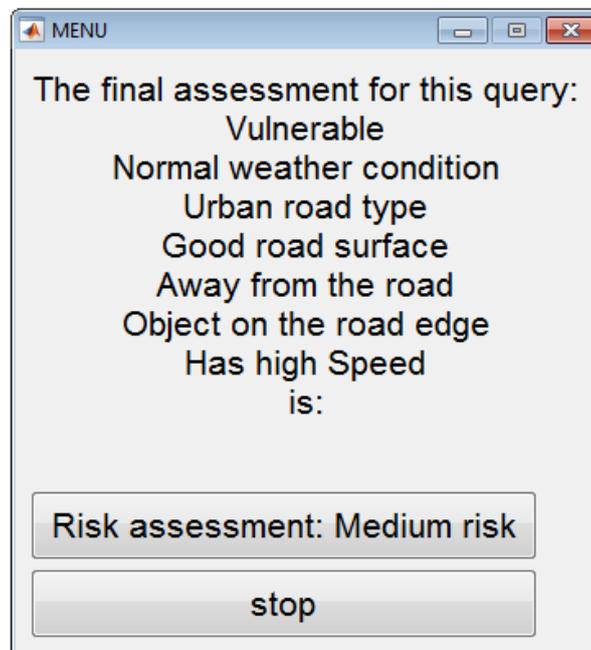


Figure B.12: MATLAB query, test rule number 10.

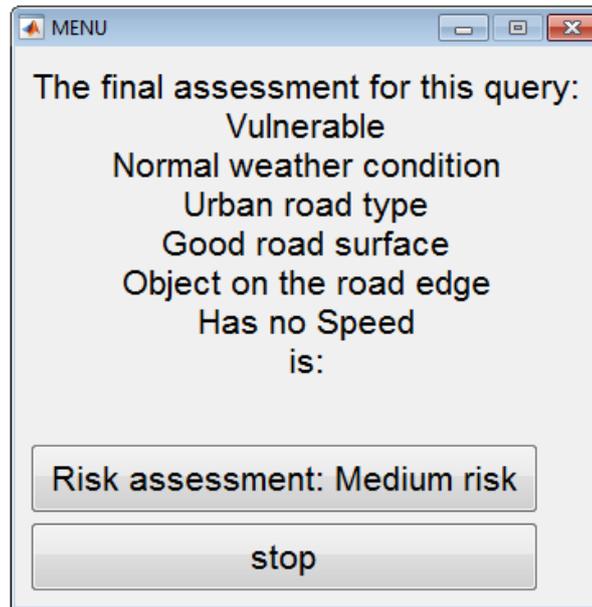


Figure B.13: MATLAB query, test rule number 11.

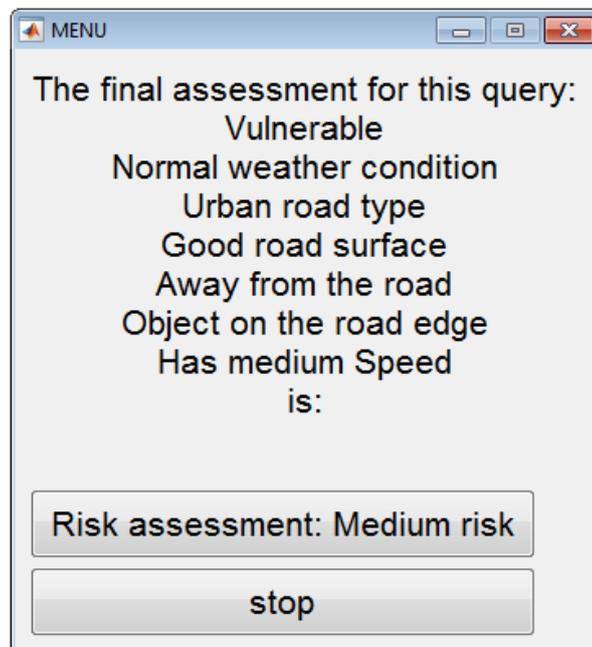


Figure B.14: MATLAB query, test rule number 12.

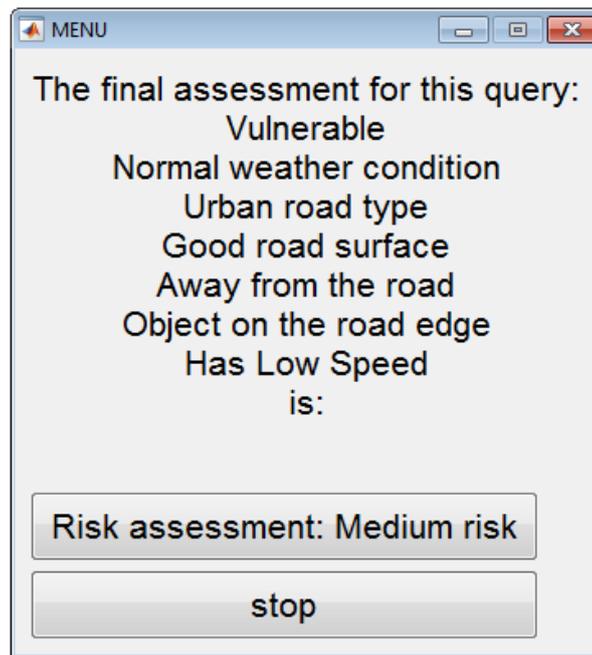


Figure B.15: MATLAB query, test rule number 13.

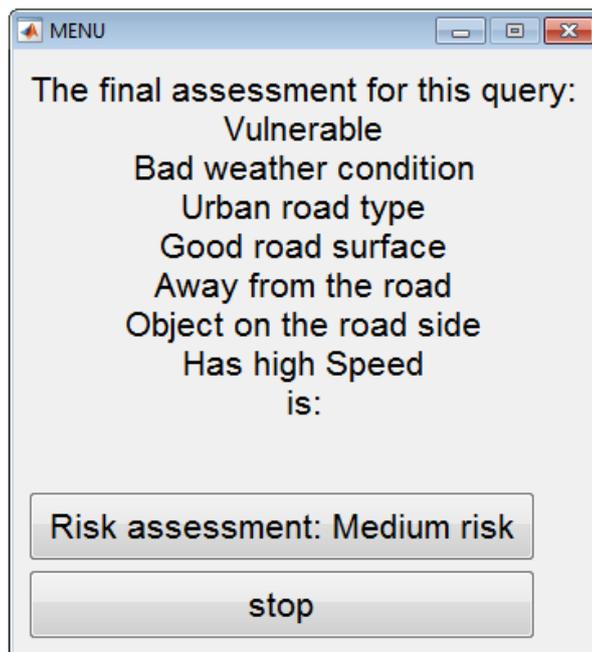


Figure B.16: MATLAB query, test rule number 14.

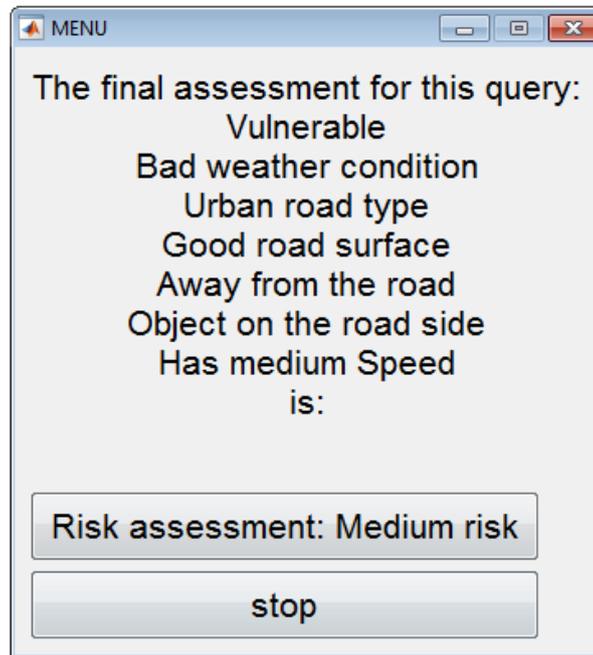


Figure B.17: MATLAB query, test rule number 15.

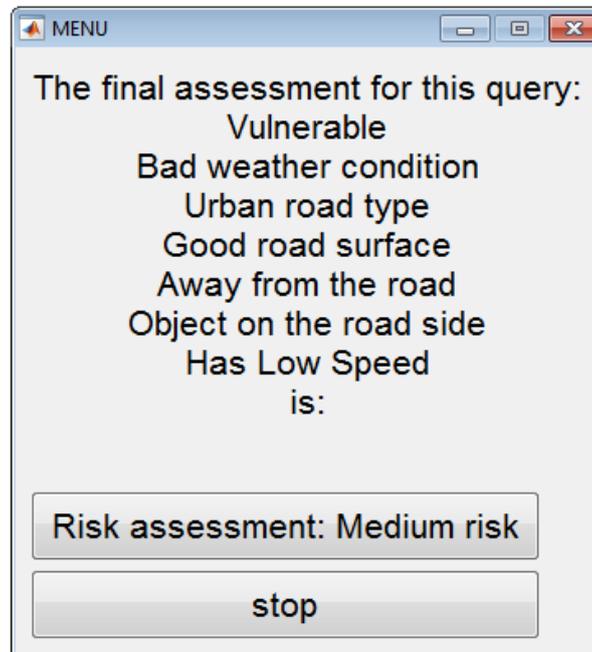


Figure B.18: MATLAB query, test rule number 16.

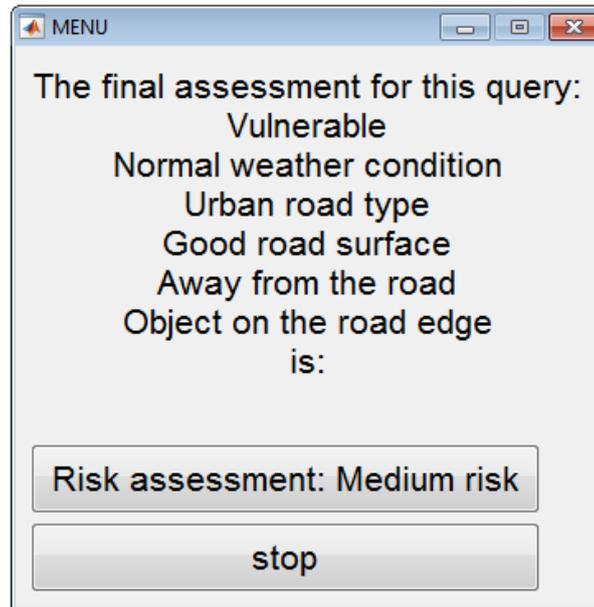


Figure B.19: MATLAB query, test rule number 17.

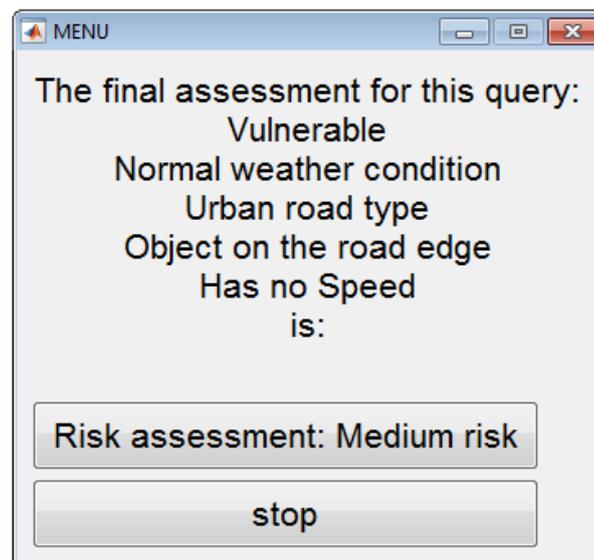


Figure B.20: MATLAB query, test rule number 18.

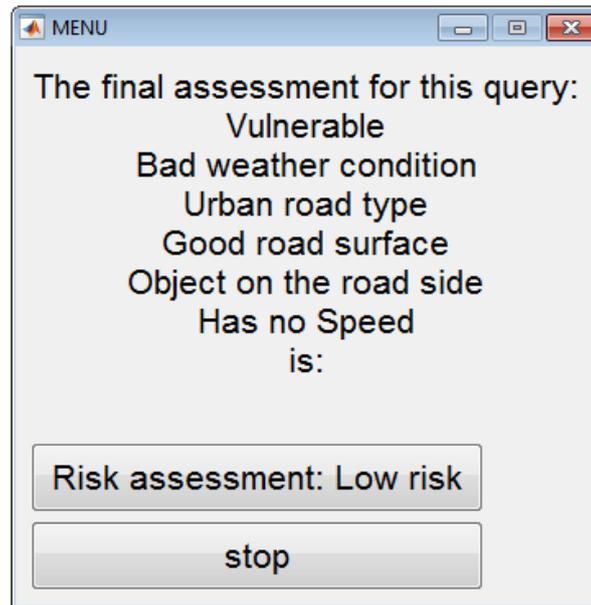


Figure B.21: MATLAB query, test rule number 19.

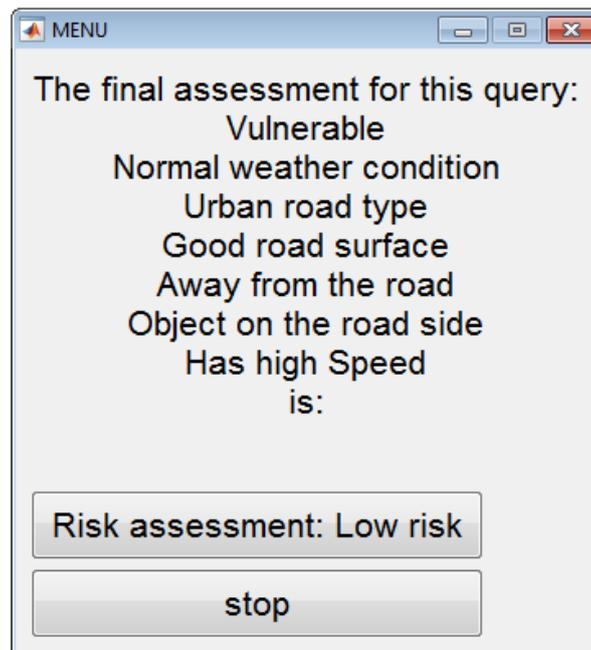


Figure B.22: MATLAB query, test rule number 20.

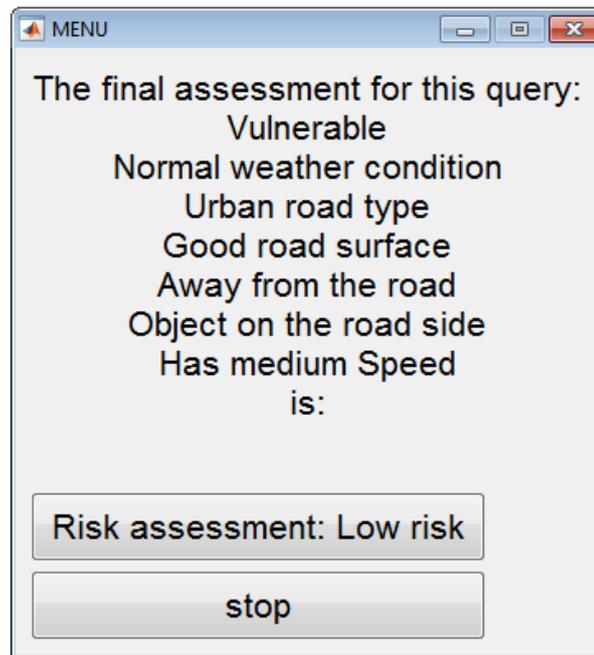


Figure B.23: MATLAB query, test rule number 21.

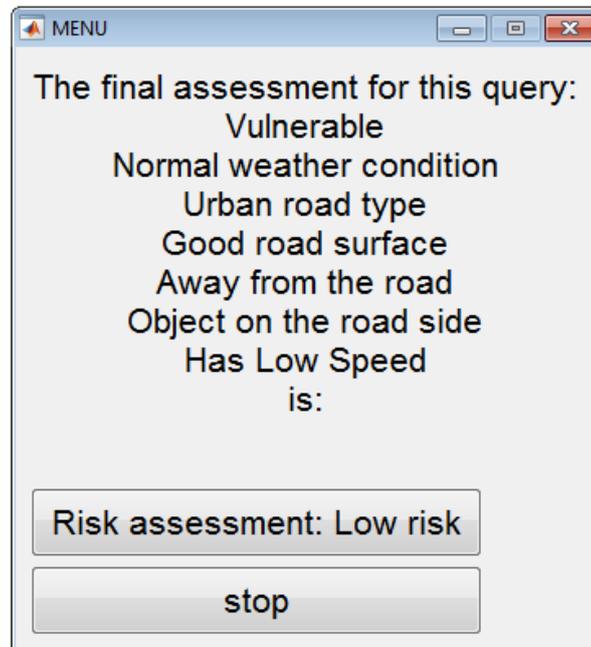


Figure B.24: MATLAB query, test rule number 22.

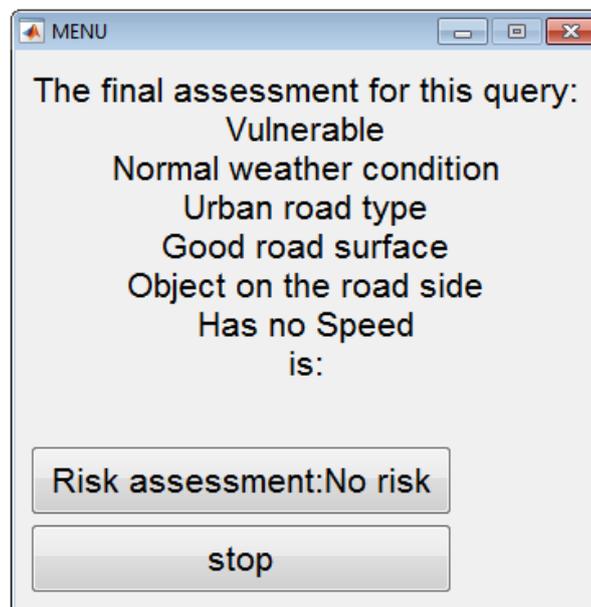


Figure B.25: MATLAB query, test rule number 23.