UNDERSTANDING OBJECT RELATIONS IN TRAFFIC SCENES

Irina Hensel, Alexander Bachmann, Britta Hummel and Quan Tran
Department of Measurement and Control, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Keywords: Intelligent vehicles, Object relations, Markov logic.

Abstract: An autonomous vehicle has to be able to perceive and understand its environment. At perception level objects are detected and classified using raw sensory data, while at situation interpretation level high-level object knowledge, like object relations, is required. In order to make a step towards bridging this gap between low-level perception and scene understanding we combine computer vision models with the probabilistic logic formalism Markov logic. The proposed approach allows for joint inference of object relations between all object pairs observed in a traffic scene, explicitly taking into account the scene context. Experimental results based on simulated data as well as on automatically segmented traffic videos from an on-board stereo camera platform are provided.

1 INTRODUCTION

To enable autonomous driving, a vehicle has to perceive and interpret its environment with respect to the driving task. Perception in this context refers to the detection and classification of objects based on raw sensory data, whereas interpretation denotes inferring and manipulating high-level scene descriptions based on this data, such as relations between objects and driver intentions. Methods from the field of *cognitive vision* aim at bridging the gap between perception and interpretation, by using abstracted sensor data together with explicitly encoded prior knowledge and inference procedures (Vernon, 2006). This prior knowledge typically takes the form of frequently occurring spatial and temporal relations between domain objects. A couple of formalisms have been applied to model and exploit such knowledge, such as: probability theory (Howarth and Buxton, 2000), frames (Hotz et al., 2008), description logic (Neumann and Möller, 2008) (Hummel et al., 2008), Situation Graph Trees (Arens et al., 2004), Scenarios (Georis et al., 2006) and lately Markov logic (Tran and Davis, 2008). Typical applications of such methods include surveillance tasks, interpretation of aerial images or analysis of traffic situations. Some related work that links traffic videos to conceptual relational knowledge is outlined in the following. (Howarth and Buxton, 2000) derive conceptual representations of events from model-based object tracking data computed on traffic videos from a roundabout. (Cohn et al., 2006) overview a system that can learn traffic behaviour using qualitative spatial relationships among close objects travelling along learned paths. Another system, presented in (Gerber and Nagel, 2008), imports extracted geometrical trajectories from inner-city monocular videos into a conceptual representation of elementary vehicle actions based on a fuzzy metric-temporal Horn logic. The same knowledge formalism is used in (Fernández et al., 2008) as a basis for an integrative architecture of a cognitive vision system, which extracts textual descriptions of a recorded pedestrian crossing scenario. (Arens et al., 2004) demonstrate that high-level hypotheses about intended vehicle behaviour can be used to improve tracking under occlusion. The next two recent works use Markov logic as a representation language. That is a novel probabilistic logic formalism, which can handle uncertain and imperfect data (Richardson and Domingos, 2006). (Tran and Davis, 2008) addresses the task of visual recognition of interactions of people and vehicles at a parking lot. (Bachmann and Lulcheva, 2009) classifies multiple independently moving objects by taking into account existing object relations.

With the exception of (Bachmann and Lulcheva, 2009) all publications mentioned use videos recorded from static cameras. Furthermore, relations between object pairs are inferred without taking scene context into account. This can lead to a globally implausible scene description since it is hard to detect noise and outliers in the sensor data.

Hensel I., Bachmann A., Hummel B. and Tran Q. (2010).
UNDERSTANDING OBJECT RELATIONS IN TRAFFIC SCENES.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 389-395
DOI: 10.5220/0002832603890395

To address these issues, this contribution presents an approach that allows for *joint* inference of relations between all object pairs in a scene, thus explicitly taking into account the scene context. Moreover, traffic videos are acquired from a stereo camera platform that is mounted inside a moving vehicle.

The proposed system automatically segments images into object hypotheses. Motion profile and position in space are estimated for every object hypothesis. This quantitative sensor data is mapped onto symbols and an evidence file is generated. Markov logic models for understanding object relations in a traffic scene are developed and trained on a set of traffic images. The evidence together with the trained model are provided as input to the Markov Logic reasoner. As a result conditional probabilities for the validity of the modelled object relations between every two objects in the scene are computed.

This paper is organised as follows: next section will give a short theoretical introduction to Markov logic. Then the vocabulary used in our traffic scene models is introduced in form of an ontology. Section 4 describes the traffic scene models developed in Markov logic. Finally, Section 5 provides experimental results on simulated and real data.

2 MARKOV LOGIC

Markov logic combines first-order logic with Markov random fields. It provides a framework for explicitly modeling relations in complex domains, while taking into account uncertainties and performing probabilistic inference (Richardson and Domingos, 2006). A Markov logic network (MLN) L consists of a set of weighted logic formulae (F_m, g_m) describing a specific domain. The formulae F_m are constructed from logical atoms (e.g. *sceneObject(o)*) linked with logical connectives and quantifiers. The attached real valued weights g_m validate the assertions stated over the domain by the corresponding formulae and can be learned from training data. Given L and a finite set of logical constants B (e.g. *O1*) all possible groundings of each logical atom X (e.g. *sceneObject(O1)*) and all possible groundings of each formula G can be instantiated by substituting all (typed) variables by constants from B. Each ground formula in G is assigned the weight of the underlying first-order formula from L. The set of ground atoms X can be seen as a set of binary random variables and therefore be represented by an Markov random fields $M(L,B)$, which has a binary node for every X_m. The value of a node is 1, if the corresponding ground atom is true and 0 otherwise. There is an edge between two nodes of $M(L,B)$ iff the corresponding ground atoms appear together in at least one element of G. Thus, all ground atoms of a ground formula constitute a clique in $M(L,B)$. The state $x_{\{m\}}$ of the m-th clique is evaluated by the feature $f_m(x_{\{m\}}) \in \{0,1\}$ of the corresponding ground formula from G and by the weight g_m assigned to it. The value of the feature $f_m(x_{\{m\}})$ is 1, if G_m is satisfied by $x_{\{m\}}$, i.e. if the ground formula is true. The joint distribution of $M(L,B)$ is

$$P(X=x) = Z^{-1}\exp\left(\sum_m g_m f_m(x_{\{m\}})\right),$$

where Z is a normalization factor. Algorithms for learning and inference in MLNs are implemented in the open-source package Alchemy (Kok et al., 2007) and have been used throughout this work.

3 ONTOLOGY

Knowledge about object attributes and object relations to other scene objects is described using a defined vocabulary. Figure 1 shows the pictorial representation of the ontology formalised in first-order logic. The arity of each predicate symbol, that is the number of its typed logical variables, is shown in brackets (as e.g. *hasSpeed(object,speed)*).

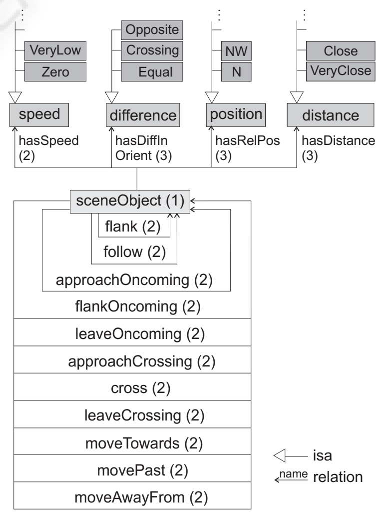

Figure 1: Object relation ontology. In order to maintain readability some conceptual values of the object attribute classes are left out. The numbers in brackets denote the arity of each predicate symbol.

Object Attributes. The scene object concept is connected with all modelled object attributes (see Figure 1). One can distinguish between self object attributes, which refer to one object, and relative

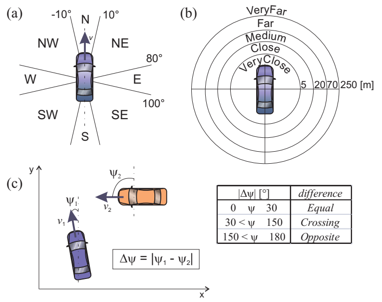

Figure 2: Conceptual values for relative position, relative distance and relative orientation.

object attributes, which refer to two objects. Respectively, there is a predicate symbol of proper arity that explicates each of these links. The quantitative value range of every modelled attribute is discretised in a proper set of conceptual values that are formalised as logical terms (e.g. *VeryLow*). The modelled object attributes are: *speed* - object speed; *difference* - difference in orientation between two objects; *position* - relative position between two objects and *distance* - relative distance between two objects. All conceptual values of an object attribute are modelled as pairwise disjoint and jointly exhaustive (see Figure 2).

Object Relations. An object relation in this work denotes an elementary action of a traffic participant supplemented by a reference to another relevant scene object. All object relations depicted in Figure 1 are formalised as predicates of arity 2 with both variables being of type *object*, as e.g. *follow(object,object)*. Every object relation is specified for the second entry (primary object) with respect to the first one (reference object), e.g. *follow(O1,O2)* reads "*O2* follows *O1*". The object relations represent general relations between two moving objects or between a moving and a standing object. The meaning of each object relation is visualised in Figure 3.

4 MODEL

This section introduces several traffic scene models developed in Markov logic. They consist of a number of first-order logic rules formulated with the predicates introduced in the previous section. These rules can be divided into hard and soft rules. Hard rules are assumed to be deterministic and obtain a large positive weight attached without a training phase. Soft rules make assertions over the domain that are only typically true. The weights associated with them are learned from hand-labeled training data generated from images of urban, rural road and highway traffic scenes.

Thereby all free logical variables in the examples below are to be considered as implicitly universally quantified.

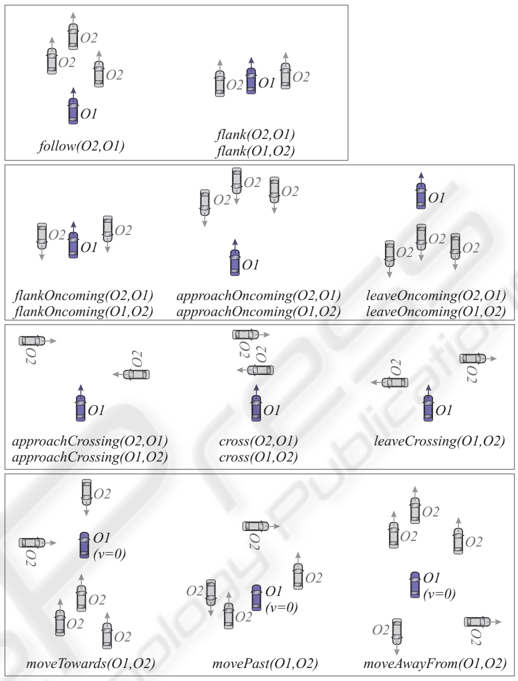

Figure 3: Exemplary traffic scenes visualising the meaning of the object relations.

4.1 Object Relations MLN (OR MLN)

Object Relations MLN (OR MLN) models dependencies between the introduced object attributes and object relations. In the training phase for this MLN a formal definition of object relations in terms of object attributes is learned.

In OR MLN hard formulae describe the taxonomical structure of the object attribute predicates and their properties, such as symmetry or disjointness. The predicates *hasRelDist* and *hasDiffInOrient* describe symmetric relative object attributes, while *hasRelPos* is unsymmetric.

Soft rules model the correspondence between object attribute values and object relations. There are rules that explicit the dependencies between the movement state of two different objects and the present object relation, e.g. if both objects are standing still none of the introduced object relations is valid, if both objects are moving none of the relations *moveTowards*, *movePast* and *moveAwayFrom* is valid, and if one of the objects is moving and the other is standing, then none of the object relations, representing occurrences between two moving objects, is valid.

Further, there are a set of rules that link the remaining three modelled object attributes with each of the object relations, e.g. :

!(o1=o2) $\wedge$ *hasRelPos(o1,o2,+p)* $\Rightarrow$ *follow(o1,o2)*
!(o1=o2) $\wedge$ *hasRelDist(o1,o2,+dist)* $\Rightarrow$ *follow(o1,o2)*
!(o1=o2) $\wedge$ *hasDiffInOrient(o1,o2,+d)* $\Rightarrow$ *follow(o1,o2)*

In the syntax of Markov logic a "!" denotes logical negation. The plus operator preceding the variables in the above example makes it possible to learn a separate weight for each formula obtained by grounding the variable with every possible conceptual value of the corresponding object attribute. This can be interpreted as learning a "soft definition" for every object relation. The weights of the soft rules are learnt using the discriminative training algorithm from the Alchemy system (Kok et al., 2007).

The full OR MLN consists of the defined hard rules and the soft rules with learned weights. It softly defines the object relations in terms of the object attributes. Figure 3 visualises several examples of these learned definitions for each object relation. Symmetric object relations are indicated in the Figure.

4.2 Scene Consistency MLN (SC MLN)

Using OR MLN one can infer the present object relations between all possible pairs of objects in a traffic scene given the object attributes. Thereby all object relations are inferred jointly. However, uncertainties in the measurement of the object attributes can still lead to a globally inconsistent scene description. This is addressed within the Scene Consistency MLN (SC MLN), which models which object relations may be valid at once among three scene objects.

SC MLN consists of soft rules constructed with object relation predicates only. Despite of rules that state which object relations are symmetrical, there are a number of rules that describe plausible object relations between three different scene objects, such as:

!(o0=o1) $\wedge$ *!(o0=o2)* $\wedge$ *!(o1=o2)* $\wedge$
follow(o1,o0) $\wedge$ *follow(o1,o2)* $\Rightarrow$
follow(o0,o2) $\vee$ *follow(o2,o0)* $\vee$ *flank(o2,o0)*

All combinations of object relations with three objects are modelled. Figure 4 shows all constructed rules in a schematic way. The abbreviations used are listed in Table 1. All of these formulae are constructed analogously to the one written above. The rows and lines in Figure 4 contain the predicates from the left side of the formula and the corresponding matrix entry contains the right side of the formula (the possible plausible object relations for this case). The formula from above, for example, is build from row one and line one.

The constructed rules are trained generatively on hand labeled training data. The weighted knowledge base forms the SC MLN.

Table 1: Abbreviations of object relation predicates used in Figure 4.

fo10	follow(o1,o0)
fl10	flank(o1,o0)
aO10	approachOncoming(o1,o0)
fO10	flankOncoming(o1,o0)
lO10	leaveOncoming(o1,o0)
aC10	approachCrossing(o1,o0)
lC10	leaveCrossing(o1,o0)
c10	cross(o1,o0)
mT20	moveTowards(o2,o0)
mP20	movePast(o2,o0)
mA20	moveAwayFrom(o2,o0)

4.3 SCOR MLN

SCOR MLN stands for the combination of the above presented SC MLN and OR MLN. It consists of all hard rules and weighted soft rules of both MLNs. While OR MLN models relations between object pairs, SC MLN models the plausibility of a scene as a whole. Thus SCOR MLN allows for a global look at a traffic scene.

4.4 Evidence/ Inference

The available quantitative information about domain objects is mapped onto logical constants using qualitative abstraction. The constants represent objects (e.g. *O1*), conceptual values of self object attributes (e.g. *Zero*) or conceptual values of relative object attributes (e.g. *NW*). The set of true ground atoms resulting from the abstracted constants (*sceneObject(O1), hasSpeed(O1,Zero)*, etc.) is the evidence given as input to the reasoner.

Based on the MLN and evidence, a grounded Markov network specifying the joint distribution is constructed and the conditional probability that a particular ground atom is true, can be inferred (e.g. that *O1* follows *O2*). This way the probability that a particular object relation holds can be estimated for every evidence object pair at every discrete time step.

5 EXPERIMENTS

Experiments were carried out on simulated data as well as on automatically segmented traffic image sequences. In all experiments the MC-SAT Algorithm

	fo12	fo21	fl12	aO12	fO12	lO12	aC12	lC21	lC12	c12	mT21	mP21	mA21
fo10	fo02,fo20,fl20	fo20	fo20	aO20	aO20	lO20,fO20,aO20	aC20,c20	aC20,c20,lC20	lC02,c20,aC20	c20,aC20,lC02	mT20,mP20	mT20	mT20,mP20,mA20
fo01	fo02	fo02,fo20,fl20	fo02	aO02,fO02,lO02	lO02	lO02	aC02,c02	lC20	c02,lC02	c02,lC02,aC02	mT20,mP20,mA20	mP20,mA20	mA20
fl01			fl02	aO02	fO02	lO02	aC02,c02,lC02	lC20	lC02,c02,aC02	lC02,aC02	mT20	mP20	mA20
aO10				fo02,fo20,fl20	fo20	fo20	c20,aC20	aC02,c02,lC02	aC20,c20,lC20	aC20,lC20	mT20,mP20,mA20	mT20	mT20
fO10					fl20	fo20	lC20	aC02,c02	*	lC20	mA20	mP20	mT20
lO10						fo02,fo20,fl20	lC20	aC20,c20,lC20	*	*	mA20	mA20	mT20,mP20,mA20
aC10							fo02,fo20,fl20,aO20	fo02,fo20,aO20,fO20	fo20,fO20,aO20,lO20	fo20,aO20	mT20	mT20	*
lC01								fo02,fo20,fl02,aO02,aC02	fo20,aO20,fO20,lO20,aC20	fo20,fl20,aO20,fO20,aC20,lC20	mT20,mP20	mT20,mP20	mT20,mP20
lC10								fo02,aO02,fO02,lO02,aC02	fo02,fo20,fl20,lO20	fo02,lO02	mA20	mA20	mA20
c10										fo20,aO20,aC20,lC02	mT20,mP20	mT20	mT20
mT20											fo01,fo10,fl01,aO01,aC01	fo01,aO01	fo10,aO01,fO01,lO01,lC01
mP20												fl01,fO01	fo10,lO01,lC01
mA20													fo01,fo10,fl01,lO01

Figure 4: SC MLN formulae encoded in a matrix form. Abbreviations are depicted in Table 1.

from Alchemy was used for inference (Richardson and Domingos, 2006).

5.1 Simulated Example

The following simulation exemplifies the need of a *global* view on a traffic scene. A situation with three cars *O0*, *O1* and *O2* is simulated. All three cars are moving in the same direction, so that for every object pair the object relation *follow* is valid. Measurement uncertainty with respect to object orientation was added and the resulting scene is shown in Figure 5. The corresponding abstracted conceptual values of all modelled object attributes are depicted in Figure 5 in brackets, while inferred results of OR MLN and SCOR MLN are shown in boxes. Thereby only the inferred object relations with highest probability are listed. Because of the simulated uncertainty for *O0*, the conceptual values for the relative object attribute difference in orientation result in *Equal* between *O0* and *O2* as well as between *O1* and *O2*, but *Crossing* between *O0* and *O1*. A contradiction in these measured attributes is easy to see, if we look at the scene described by the object attributes as a whole: if *O2* has qualitative the same orientation as *O1* and *O0* has crossing orientation to *O1*, then the orientation between *O0* and *O2* should be crossing too; or, if same orientation between *O1* and *O2* and between *O0* and *O2* holds, then it should hold also between *O0* and *O1*. The OR MLN is not capable of resolving this contradiction, since it infers the object relations considering the object attributes only. The SCOR MLN, however, takes the consistency of the scene into account, which leads to a considerable increase in the conditional probability for *follow* from 0.02 to 0.44.

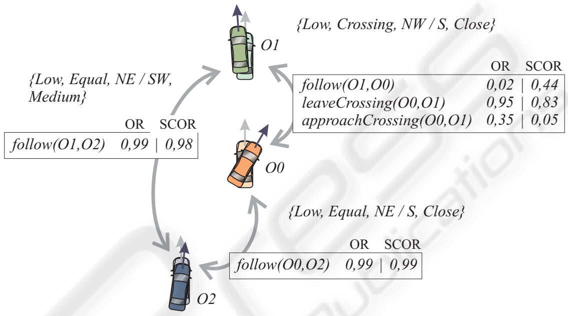

Figure 5: Simulated traffic scene example with corresponding object attribute values and inference results of OR MLN and SCOR MLN.

5.2 Real Data

Video data from an on-board stereo camera platform are processed with the algorithm described in (Bachmann and Dang, 2008) to automatically segment and track object hypotheses. This method partitions the image sequence into independently moving regions with similar 3-dimensional motion and relative distance. For every tracked segmented object hypothesis we obtain a unique identifier. As long as a particular segmented hypothesis is being tracked, we get for each frame a bounding box with its dimensions and height above the estimated ground plane, the characteristic 3D motion of the corresponding region and the current position in space. These quantitative measurement series are subsequently preprocessed and then mapped onto conceptual values.

The series preprocessing step is done in batch mode for a segmented image sequence. At first the measurement series are smoothed. Afterwards object speed magnitude and direction for every frame are calculated from the corresponding 3D motion profile. Further, relative distance, relative position and difference of orientation are computed for each possible pair of moving objects in the frame. Thereby we consider the ego-vehicle as a scene object, so that pair relations between the segmented object hypotheses and the ego-vehicle are also evaluated. The difference in orientation is determined by substracting the speed direction angles of both objects. By computing the relative position between a reference and a primary object a reference system centered at the reference object is used. The reference axis is thereby the axis in the direction of motion. We compute the relative position relation exhaustively for every possible combination of reference and primary object. The

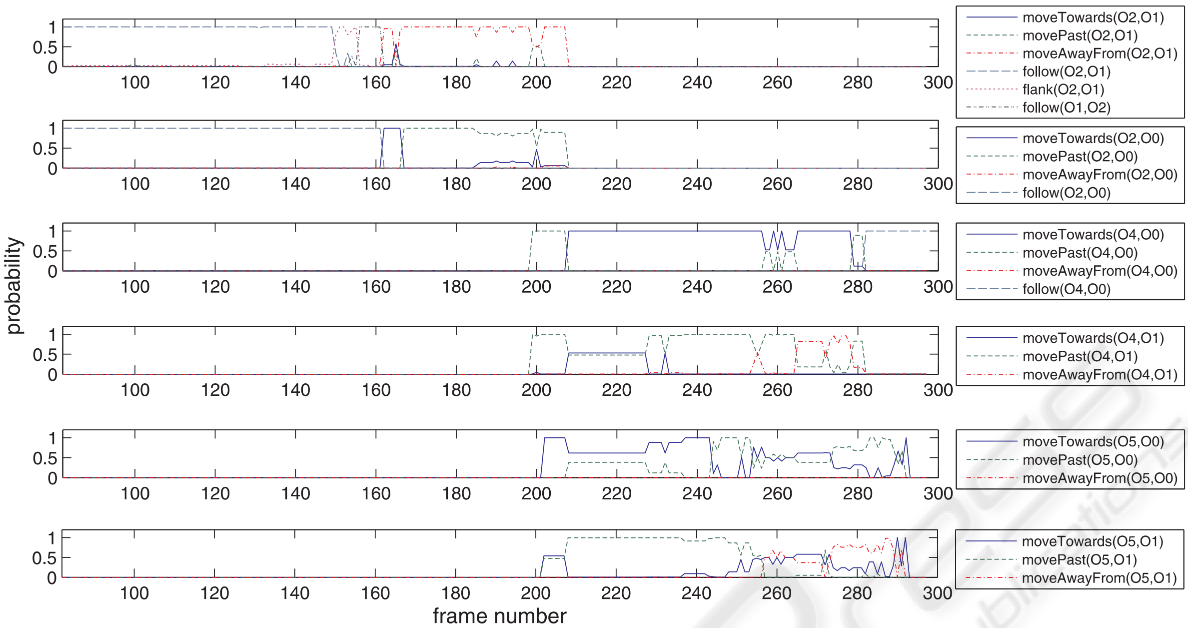

Figure 6: Inference results for selected groundings of the query predicates for all frames of the inner-city test sequence obtained using SCOR MLN.

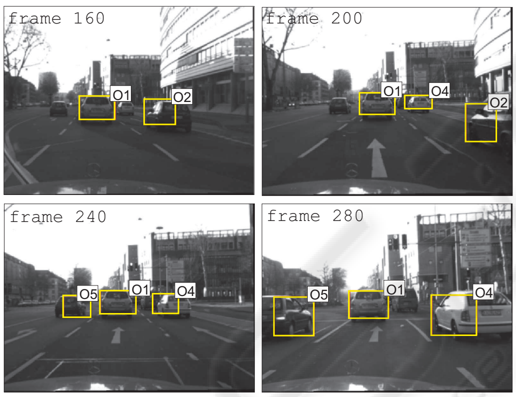

Figure 7: Representative frames from the inner-city test sequence with segmented object hypotheses.

calculated quantitative values for all necessary object attributes and relations are adjacently represented by conceptual values and abstracted to ground logical atoms. This abstraction step is carried out for each frame of the sequence and thus we get as a result one evidence file per frame. Inference is run for every evidence file generated and so we get inferred probabilities for each frame of the corresponding sequence. Hence, one should consider the results obtained for each frame as an individual experiment, which can be assessed as being acceptable or not. Experiments on automatically segmented traffic video sequences are performed with both OR MLN and SCOR MLN. Thereby query predicates are all modelled object relations. Exemplary results of SCOR MLN for an inner-city video sequence are visualised and discussed in the following. Figure 7 show several characteristic image frames from the test sequence that reveal the temporal traffic activities. Object IDs and bounding boxes of the segmented objects in these images are depicted too. In Figure 6 the results for all frames of the test sequence are represented as graphs of inferred probability versus image frame number for selected groundings of the query predicates. It can be seen that the inferred results comply to a great extend with the sequence ground truth: First, the ego-vehicle *O0* follows *O1* and *O2*; then, *O2* stops so that *O1* and *O2* drive by; while *O2* disappears from camera sight after some time, *O4* and *O5* appear standing still, waiting at the traffic light; eventually *O1* drives between *O4* and *O5*; the ego-vehicle *O0* follows *O1* throughout the sequence.

Table 2: AUC ROC results for OR MLN and SCOR MLN.

	OR MLN	SCOR MLN
follow	0,994051	0,997625
flank	0,574146	0,820281
moveTowards	0,929208	0,919756
movePast	0,900425	0,951633
moveAwayFrom	0,962180	0,978520

In order to gain a quantitative measure for the accuracy of the inferred probabilities, our approach can be considered as a classification task. Therefore, the area under the receiver operating curve (AUC ROC) is computed for each object relation seen as a different class. Ground truth for the validity of the object relations in every frame from the test sequence is manually annotated. It should be noted that this is an ambiguous task depending on the judgement of the human observer. The AUC ROC results of OR MLN and SCOR MLN for the test sequence visualised in Figure 7 are listed in Table 2. Hereby the inferred probabilities of 3384 groundings per modelled object relation were evaluated. The SCOR MLN achieved significantly improved results for most relations, supporting the claim for considering scene context in complex relational classification tasks.

6 CONCLUSIONS

This contribution introduced an approach for inferring a conceptual representation of relations between objects in traffic scenes using Markov logic. Soft definitions for object relations in terms of discretised sensor data were learned, as well as typical combinations of such object relations. These learned models were tested on automatically segmented traffic videos from an on-board stereo camera platform. Taking into account both the soft definitions and typical scene context, the conditional probability of several object relations given the learned model and evidence was computed for each object pair in each frame of a test sequence. The results complied in most cases with the judgement of a human observer. The proposed approach can be seen as a promising step towards bridging the gap between low-level image processing and high-level situation interpretation. Future work considers verifying the proposed approach on a broader statistical base, augmenting the model with temporal dependencies and closing the loop to low-level scene segmentation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge support of this work by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the Transregional Collaborative Research Centre 28 “Cognitive Automobiles”.

REFERENCES

Arens, M., Ottlik, A., and Nagel, H. H. (2004). Using behavioral knowledge for situated prediction of movements. In *KI*, pages 141–155. Springer-Verlag.

Bachmann, A. and Dang, T. (2008). Improving motion-based object detection by incorporating object-specific knowledge. *International Journal of Intelligent Information and Database Systems (IJIIDS)*, 2(2):258–276.

Bachmann, A. and Lulcheva, I. (2009). Combining low-level segmentation with relational classification. In *ICCV2009; IEEE Workshop on Visual Surveillance (VS)*, pages 1216–1221.

Cohn, A. G., Hogg, D., Bennett, B., Devin, V., Galata, A., Magee, D., Needham, C., and Santos, P. (2006). Cognitive vision: Integrating symbolic qualitative representations with computer vision. In Christensen, H. I. and Nagel, H. H., editors, *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, volume 3948 of *LNCS*, pages 221–246. Springer.

Fernández, C., Baiget, P., Roca, X., and Gonzílez, J. (2008). Interpretation of complex situations in a semantic-based surveillance framework. *Image Commun.*, 23(7):554–569.

Georis, B., Mazire, M., Brémond, F., and Thonnat, M. (2006). Evaluation and knowledge representation formalisms to improve video understanding. In *Proceedings of the International Conference on Computer Vision Systems (ICVS'06)*, New-York, NY, USA.

Gerber, R. and Nagel, H. H. (2008). Representation of occurrences for road vehicle traffic. *Artif. Intell.*, 172(4-5):351–391.

Hotz, L., Neumann, B., and Terzic, K. (2008). High-level expectations for low-level image processing. In *KI*, pages 87–94. Springer-Verlag.

Howarth, R. J. and Buxton, H. (2000). Conceptual descriptions from monitoring and watching image sequences. *Image Vision Comput.*, 18(2):105–135.

Hummel, B., Thiemann, W., and Lulcheva, I. (2008). Scene understanding of urban road intersections with description logic. In Cohn, A. G., Hogg, D. C., Möller, R., and Neumann, B., editors, *Logic and Probability for Scene Interpretation*, number 08091 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany.

Kok, S., Sumner, M., Richardson, M., Singla, P., Poon, H., Lowd, D., and Domingos, P. (2007). The Alchemy system for statistical relational AI. Technical report, Department of Computer Science and Engineering, University of Washington, Seattle, WA.

Neumann, B. and Möller, R. (2008). On scene interpretation with description logics. *Image Vision Comput.*, 26(1):82–101.

Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine Learning*, 62(1-2):107–136.

Tran, S. D. and Davis, L. S. (2008). Event modeling and recognition using markov logic networks. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 610–623, Berlin, Heidelberg. Springer-Verlag.

Vernon, D. (2006). The space of cognitive vision. In Christensen, H. I. and Nagel, H. H., editors, *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, volume 3948 of *LNCS*, pages 7–26. Springer.