ROAD INTERPRETATION FOR DRIVER ASSISTANCE BASED ON AN EARLY COGNITIVE VISION SYSTEM

Emre Başeski[†], Lars Baunegaard With Jensen[†], Nicolas Pugeault[†], Florian Pilz[±]
Karl Pauwels[‡], Marc M. Van Hulle[‡], Florentin Wörgötter[*] and Norbert Krüger[†]

[†] *The Mærsk Mc-Kinney Møller Institute, Univeristy of Southern Denmark, Odense, Denmark*

[±] *Department for Media Technology, Aalborg University Copenhagen, Copenhagen, Denmark*

[‡] *Laboratorium voor Neuro- en Psychofysiologie, K.U.Leuven, Belgium*

[*] *Bernstein Center for Computational Neuroscience, University of Göttingen, Göttingen, Germany*

Keywords: Large scale maps, Lane detection, Independently moving objects.

Abstract: In this work, we address the problem of road interpretation for driver assistance based on an early cognitive vision system. The structure of a road and the relevant traffic are interpreted in terms of ego-motion estimation of the car, independently moving objects on the road, lane markers and large scale maps of the road. We make use of temporal and spatial disambiguation mechanisms to increase the reliability of visually extracted 2D and 3D information. This information is then used to interpret the layout of the road by using lane markers that are detected via Bayesian reasoning. We also estimate the ego-motion of the car which is used to create large scale maps of the road and also to detect independently moving objects. Sample results for the presented algorithms are shown on a stereo image sequence, that has been collected from a structured road.

1 INTRODUCTION

A driver assistance system or an autonomous vehicle requires a road interpretation in terms of layout of the road and the relevant traffic. The main elements of such a road interpretation can be summarized as a) the lanes of the road, b) the rigid body motion estimation of the car, c) independently moving objects (IMOs) on the road and d) large scale maps of the road. In this article, we present algorithms for all the sub-problems, based on information provided by an early cognitive vision system (Krüger et al., 2004). We make use of visually extracted data from a stereo camera system. A hierarchical visual representation is created to interpret the whole scene in 2D and 3D. We then make use of temporal and spatial disambiguation mechanisms to increase the reliability of the extracted information. The layout of the street is calculated from the disambiguated data by using Bayesian reasoning. The ego-motion estimation of the car is done within the same representation and this information is used to create large scale semantic maps of the road as well as to detect and interpret the IMOs. In Figure 1,
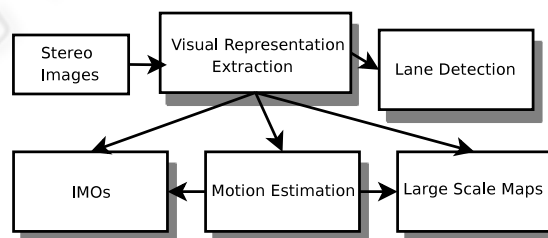

Figure 1: General overview of the system.

the general overview of the extraction process of the whole interpretation is presented. Note that, all sub-parts of the system use the same visual representation based on an early cognitive vision system.

Lane detection and parametrization, in particular in 3D, is acknowledged to be a difficult problem on which a lot of research is performed currently (Bertozzi and Broggi, 1998; Wang et al., 2004; Mc-Call and Trivedi, 2004). The difficulty comes from a) the extraction of relevant information in far distances, b) the extraction of 3D information (in particular in hilly environments), and c) poor image quality and highly varying lightning conditions (leading,

Figure 2: Satellite image of the road from which the reference stereo image sequence has been collected. Note that, the driving direction is from left to right.

e.g., to cast shadows). For 2D and 3D lane interpretation, there are already systems build into cars (see (Bertozzi et al., 2000) for an overview) which however only give information at the current position while for driving future positions and other parameters (such as curvature) are of importance.

Moreover, besides the interpretation of the lane itself, the motion and shape of independently moving objects (IMOs) need to be integrated into such lane descriptions which, especially in 3D, is a very hard problem since not only the scene needs to be segmented into IMOs and non-IMOs but also the 6D motion vector as well as the shape of the IMOs need to be computed based on only the small image areas the IMOs appear in. Current approaches towards this problem typically integrate object detection, tracking, and geometrical constraints (see e.g., (Leibe et al., 2007)). Such methods crucially rely on object detection, which is very difficult for small or distant objects. Our method on the other hand, can provide the required higher precision by integrating multiple dense cues, namely dense disparity, dense optical flow, and self-motion derived from this optical flow. The methods used to obtain these cues (see e.g., (Pauwels and Van Hulle, 2008)) are robust to typical nuisance factors observed in real-world video sequences recorded by moving observers, such as contrast changes, unstable camera motion, and ambiguous camera motion configurations.

The ego-motion estimation of the car is not only crucial for extraction of IMOs but also for creating 3D maps of the road. These maps are similar to SLAM (see e.g., (Lemaire et al., 2007)) applications and they can be used to generate global driving events like the car's trajectory. Also, this kind of maps are important in terms of learning the behavior of the driver to use for driver assistance.

The extraction of such complex information is time-consuming. In our approach, real-time is not thought to be achieved through simplification of processes via specializing on sub-aspects, but by parallel processing on hardware. In this context, we have developed a hybrid architecture similar to (Jensen et al., 2008), consisting of different FPGAs as well as a coarse parallel computer.

The results explained in this article are also to be seen as a first step towards reasonable benchmarking in the driver assistance domain. It has been discussed in, e.g., (Hermann and Klette, 2008; Klette, 2008), that the currently used benchmarks in the computer vision community are very suitable to evaluate stereo and optic flow algorithms in controlled environments but are not very suitable to evaluate algorithms in the context of driver assistance systems since the demands are very different from the existing benchmarks. On the other hand, Klette et al. is in the process of establishing appropriate standard sequences of realistic complexity that are published on 'http://www.citr.auckland.ac.nz/6D/' where also the sequence on which we show results in this paper is available. In this way, we think this paper fosters the process of comparing results on realistic databases for driver assistance.

The rest of the article is organized as follows: In Section 2, the early cognitive vision system that has been used for this work is explained briefly. The algorithms for the sub-parts of the road interpretation _motion estimation, large scale maps, lane detection and independently moving objects_ are presented in Sections 3, 5, 6 and 7 respectively while the disambiguation process is discussed in Section 4. Note that, for evaluation of the presented algorithms, sample results are demonstrated on a reference sequence of stereo images collected from a driving on the road that is shown in Figure 2. We conclude with a discussion on these sample results in Section 8.

2 VISUAL REPRESENTATIONS

In this work we use a representation of visual information based on local descriptors called 'primitives' (Krüger et al., 2004). Primitives are extracted sparsely along image contours, and form a feature vector aggregating multiple visual modalities such as position, orientation, phase, color and optical flow. This results in the following feature vector:

$$\pi = (\mathbf{x}, \theta, \phi, (\mathbf{c_l}, \mathbf{c_m}, \mathbf{c_r}), \mathbf{f}) \qquad (1)$$

Image contours are therefore encoded as collinear strings of primitives. Because of that, collinear and

similar primitives are denoted as 'groups' in the following. These are matched across two stereo views, and pairs of corresponding 2D-primitives afford the reconstruction of a 3-dimensional equivalent called 3D–primitive, and encoded by the vector:

$$\Pi = (\mathbf{X}, \Theta, \Phi, (\mathbf{C_l}, \mathbf{C_m}, \mathbf{C_r})) \qquad (2)$$

Moreover, if two primitives are collinear and similar in an image, and their correspondence in the second image are also, then the two reconstructed 3D-primitives are grouped. Extracted 2D and 3D primitives for a sample stereo image pair are illustrated in Figure 3.

3 MOTION ESTIMATION

In the driving context, rigid body motion is prevalent. In our work, the motion is estimated using a combination of SIFT (Lowe, 2004) and primitives features (see Section 2) in a RANSAC (Fischler and Bolles, 1981) scheme. SIFT features are scale invariant features that have been very successfully applied to a variety of problems during the recent years. They afford to be matched very robustly across different views, and provide a robust basis to the motion estimation scheme. On the other hand, their localisation is imprecise. The primitives provide very accurate (sub–pixel) localisation, but are less readily matched. Moreover, because the primitives are local contour descriptors they only provide motion information in the direction normal to the contour's local orientation. This is an issue in a driving scenario where the main features, the road lines, are mostly radially distributed from the car's heading direction. This means that most of the extracted primitives yield very little information about translation in the z axis. This limitation can be remediated by using a mixture of features for motion estimation. In this work, we integrated both features in a RANSAC scheme, where SIFT are used for the initial estimation, and the consensus are formed using a combination of SIFT and primitives. For the motion estimation algorithm we chose (Rosenhahn et al., 2001), since in addition of being able to deal with different visual entities, it does optimization using a twist formulation which directly acts on the parameters of rigid-body motions.

The results presented in this chapter contain motion estimation results for feature sets consisting of primitives, SIFT and a combination of both. Results are shown in Figure 4, where each column depicts the translational and rotational motion components for one type of feature set.

The first row in Figure 4 depicting the translation, the z-axis corresponds to the car's forward mo-

tion. Here, results show that SIFT and the combination of SIFT and primitives provide much more stable results. However outliers still remain, caused by speed-bumps, potholes etc. These results in blurred images as on frame 720-730 caused by a bump on the bridge, making matching and motion estimation a difficult task.

The second row in Figure 4 depicts the rotational component where y-axis corresponds to rotation caused by steering input. The rotation results for the y-axis using SIFT and the combination of SIFT and primitives nicely corresponds to the satellite image presented in Figure 2. This correspondence between sub-parts of the road and sub-parts of the motion estimation plot is shown in Figure 10 (b). Figure 5 shows translation and rotation obtained after applying a Bootstrap Filter (Gordon et al., 1993) (using 1000 particles). This results in the elimination of the largest variations in the estimated motions and to more stable results.

4 DISAMBIGUATION

Since 3D-primitives are reconstructed from stereo, they suffer from noise and ambiguity. Noise is due to the relatively large distance to the objects observed, and the relatively small baseline of the stereo rig (33 cm). The ambiguity rises from the matching problem: despite their multi-modal aspect, the primitives only describe a very small area of the image, and similar primitives abound in an image. The epipolar constraint limits the matching problem, yet it is unavoidable that some ambiguous stereo matches occur (Faugeras, 1993). We introduce two means of disambiguation making use of temporal (Section 4.1) and spatial (Section 4.2) regularities employed by the early cognitive vision system.

4.1 Temporal Disambiguation

A first means of disambiguation is to track primitives over time. We perform this tracking in the 3D space, to reduce the likelihood of false positives. This involves resolving three problems: first, estimating the motion; second, matching predicted and observed primitives; and third, accumulating representation over time.

Using the 3D-primitives extracted at a time t and the computed ego–motion of the car between times t and $t + \delta t$, we can generate predictions for the visual representation at time $t + \delta t$. Moreover, conflicting hypotheses (reconstructed from ambiguous stereo matches) will generate distinct predictions. In most
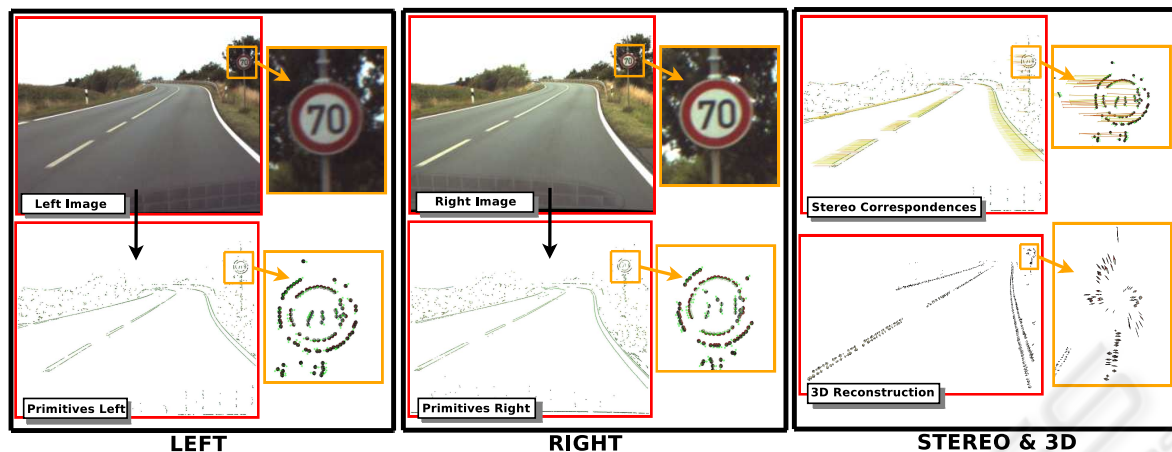
Figure 3: Extracted 2D and 3D primitives for a sample image pair. Note that, 2D primitives are used to reconstruct 3D primitives.
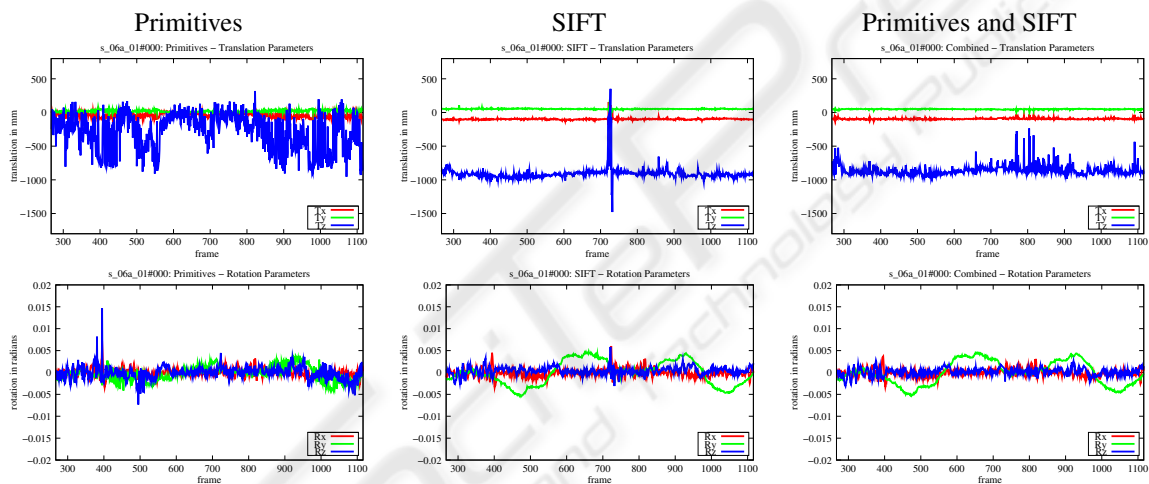

Figure 4: Motion estimation results for different feature sets over the entire stereo image sequence (frame 267-1116).

cases, only one of these predictions will be confirmed by the visual representation observed at instant $t + \delta t$. This means that, over a few frames, stereo correspondences can be disambiguated by tracking. We combined the tracking information into one accumulated likelihood measure that we use to discard unlikely hypotheses (Pugeault et al., 2008). On the other hand, primitives that achieve a high enough likelihood are flagged as correct, and preserved even if they thereafter leave the field of view or become occluded.

4.2 Spatial Disambiguation via NURBS Parametrization

The second disambiguation scheme employed here is using the collinear groups formed in both images. Stereo matches that do not agree with the group struc-

ture of a primitive are discarded, leading to a reduction of ambiguity (Pugeault et al., 2006).

Moreover, it is advantageous to have an explicit parameterization of collinear groups to allow for a controlled and more condensed description of such entities as well as for having a good control over parameters such as position, orientation and curvature along these structures. NURBS (Non-uniform Rational B-Splines) (Piegl and Tiller, 1995) has been chosen as a suitable mathematical framework since it is invariant under affine as well as perspective transformations, can be handled efficiently in terms of memory and computational power and offer one common mathematical form for both standard analytical shapes and free-form shapes. Note that, these properties hold true for 3D as well as 2D data.

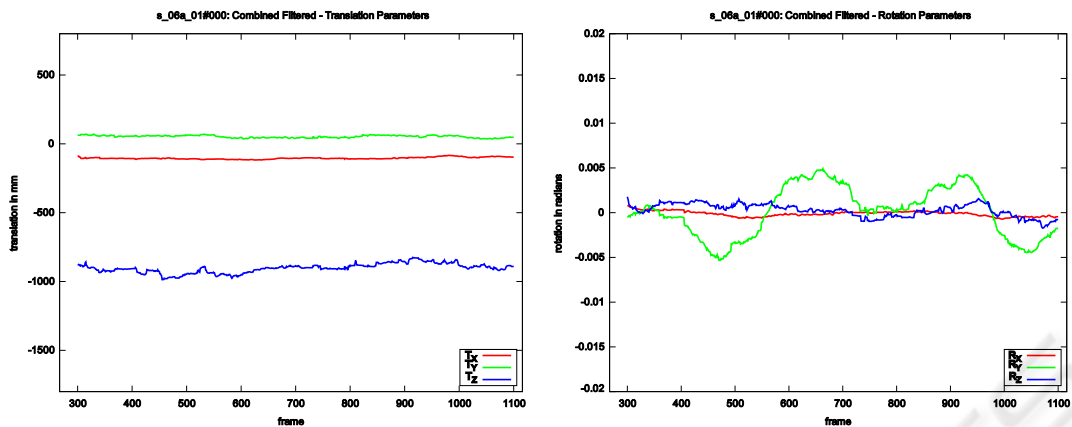NURBS parametrization is not used only for an-

Figure 5: Motion estimation results after applying Bootstrap filter to results in Figure 4.
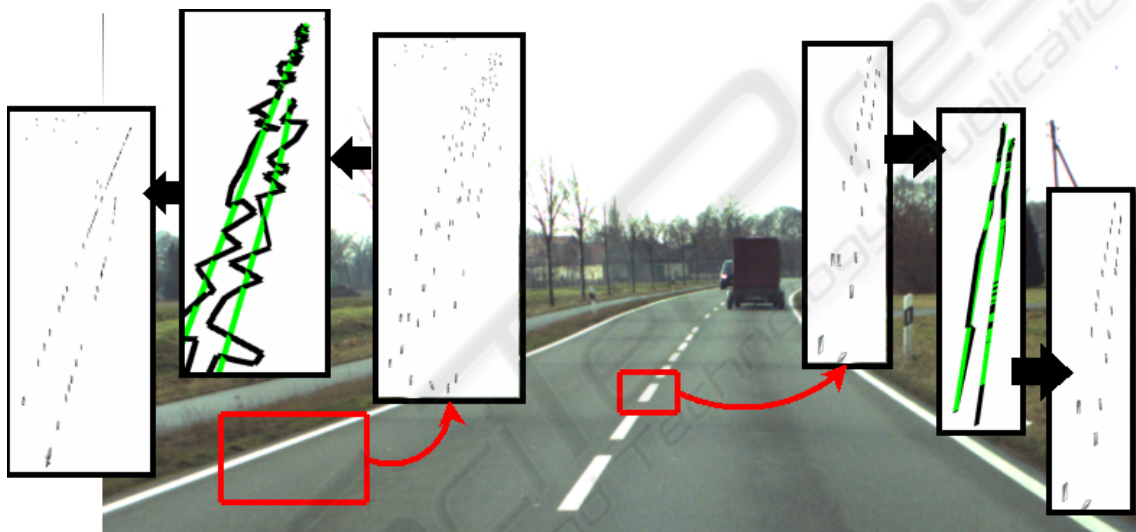

Figure 6: Position and orientation correction of 3D features by using NURBS. Green lines represent NURBS and black lines represent groups of features.

alytical calculations that are necessary to create the lane parametrization described in Section 6, but also used for disambiguation in terms of position and orientation correction of 3D features. The correction procedure is illustrated in Figure 6. After fitting NURBS (represented as green lines) to groups of 3D features (represented as black lines), position and orientation of each feature is recalculated. The procedure is shown on a good reconstruction (middle road marker) as well as a bad one (left lane marker).

5 LARGE SCALE SEMANTIC MAPS

One additional information provided by tracking is that it allows us to identify new information that is not tracked yet. All observed 3D-primitives that are not matched with one pre-existing, tracked 3D-primitive are added to the accumulated representation. This allows for integrating visual information over larger scales (Pugeault et al., 2008), in a manner similar to SLAM applications. The difference being here that the features densely describe the shape of the visual world, rather than being mere clouds of landmarks. This allows, over a scale of a few seconds for generating a richer representation of more global driving events, like road curvature, and the car's trajectory. The accumulated representation on the driving sequence in Figure 2 is shown in Figure 7, from different perspectives and at different points of the road. The first (top-left) image shows the whole road segment (in perspective).
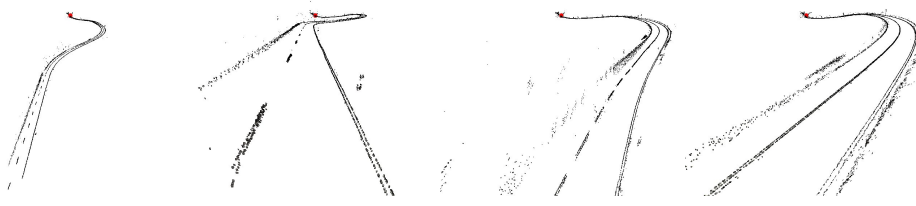
Figure 7: Large scale accumulation: four different viewpoints on the accumulated maps after 1000 frames (40s.).

6 BAYESIAN LANE FINDING AND PARAMETRIZATION

In this section, a framework for finding road lane markers based on Bayesian reasoning and parametrization of the lanes in terms of feature vectors are discussed. The visual representation discussed in Section 2 does not only provide local image features but also provides relations between these features (Başeski et al., 2007) (e.g., proximity, cocolority, distance). We want to find the primitives constituting the lane in a process combining the different relations.

To merge the different cues/relations as well as to deal with the uncertainties present in a vision system, we make use of a Bayesian framework. The advantage of Bayesian reasoning is that it allows us to:

- make explicit statements about the relevance of properties/relations for a certain object,

- introduce learning in terms of prior and conditional probabilities, and

- assess the relative importance of each type of relation for the detection of a given object, using the prior probabilities.

Let $P(e_i^\Pi)$ be the prior probability of the occurrence of an event e_i^Π (e.g., the probability that a primitive lies in the ground plane, or that two primitives are coplanar, collinear or have a certain distance to each other).Then, $P(e_i^\Pi | \Pi \in O)$ is the conditional probability of the visual event e_i given an object O.

Our aim is to compute the likelihood of a primitive Π being part of an object O given a number of visual events relating to the primitive:

$$P(\Pi \in O | e_i^\Pi). \qquad (3)$$

According to Bayes formula, equation 3 can be expanded to:

$$\frac{P(e_i^\Pi | \Pi \in O)P(\Pi \in O)}{P(e_i^\Pi | \Pi \in O)P(\Pi \in O) + P(e_i^\Pi | \Pi \notin O)P(\Pi \notin O)}. \qquad (4)$$

Figure 9: Extracted right lanes (shown in red) for some images in the reference sequence.

Using this framework for detecting lanes, we first need to compute prior probabilities. This is done by hand selecting the 3D primitives being part of a lane in a range of scenes and calculating the different relations and attributes for these selections. The resulting probabilities will reveal which relations and attributes are relevant for detecting this object.

Figure 8 shows the results of using the Bayesian framework with the relevant relations for detecting lane markers in an outdoor scene.

After finding the 3D features that potentially belong to the road by using Bayesian reasoning, the right lane of the road is extracted by using small heuristics in 2D. Since the lane starts as two separate stripes and becomes a single line further from the camera, we eliminate the short groups that are neighbor to a long group if the color distribution between the two groups is uniform. After eliminating the short groups, we select the group closest to the bottom-right corner of the image as the right lane. Note that, although the potential street lanes were found in 3D via Bayesian reasoning, the hierarchy in the representation allows us to do reasoning in 2D for the last step by using the link between 3D and 2D groups. In Figure 9, some results that show the right lane of the road are presented. Once the right lane of the road is de-

501

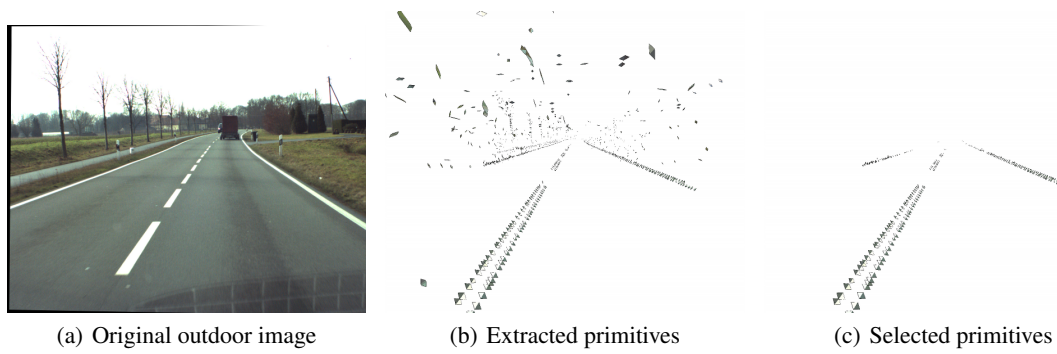(a) Original outdoor image   (b) Extracted primitives   (c) Selected primitives

Figure 8: Extracting the lane in an outdoor scenario

tected, it is necessary to create a parametrization of the lane so that it can be used for driver assistance. The idea is to represent certain locations of the lane as a feature vector, depending on the dimension of the data. In 2D, the feature vector is $v_{y_l} = [x_l, \kappa_l, s_l]$ where x_l is the x coordinate corresponding to a fixed y coordinate y_l on the lane and κ_l, s_l are the curvature and slope at point (x_l, y_l) respectively. Note that, for a curve given parametrically as $c(t) = (x(t), y(t))$, the curvature is defined as:

$$\kappa = \frac{|x' y'' - y' x''|}{(\sqrt{x'^2 + y'^2})^3} \tag{5}$$

In 3D, the feature vector is $V_{Z_l} = [X_l, Y_l, i_l, j_l, k_l]$ where (X_l, Y_l) is the x and y coordinates corresponding to a fixed Z coordinate Z_l on the lane and $[i_l, j_l, k_l]$ is the tangent vector at point (X_l, Y_l, Z_l). Note that, fitting NURBS to the right lane plays an important role in these calculations since it allows applying similar procedures for both 2D and 3D data. Also, analytical calculations like derivation is more stable and easy to compute when NURBS is used.

In Figure 10(d-f), the 2D lane parameter extraction results are presented for the reference sequence collected from the road shown in Figure 2. For each image in the sequence, a 2D feature vector has been found from the right lane for y coordinate equals to 500 pixels. The road has two sharp right turns and one less sharp left turn with the order right, left, right. The plots regarding the slope and x coordinate show all three turns as shown in Figure 10 (d) and (e). On the other hand, the plot for curvature (Figure 10 (f)) can show only two sharp right turns and gives slight indication of left turn. Position and tangent vector describe the structure of the road in 3D as well, as shown in Figure 10 (a) and (c).

7 INDEPENDENTLY MOVING OBJECTS

One important source of information in a driving context lies with the other vehicles (and more generally all other moving objects), their detection, tracking and identification. In this work, Independently Moving Objects (IMOs) are detected by subtracting the optic flow introduced by the car's self-motion from the observed optical flow as described in (Pauwels, 2008). This provides a list of IMOs, with appropriate bounding boxes (BB). We use this information to segment the image in areas (background and IMOs) where visual information is extracted and accumulated separately (as in Section 4.1).

Figure 11 illustrates the IMO handling module of the system. In (a) the detected IMOs are identified by bounding boxes and an unique ID; (b) the left image patch contained in the IMO's BB is extracted and magnified to increase resolution; (c) the right BB's location is estimated using the mean disparity of the primitives in the left BB, and the corresponding patch in the right image is also extracted and magnified; (d) new primitives are extracted on the magnified images and matched across both patches. The stereo matches are used for reconstructing 3D-primitives, as before, and accumulated over time.

Figure 12 shows the segmentation of the IMOs. On the left, the figure shows a detail of the accumulated road. The IMO has been segmented out. On the right the accumulated representation of the IMO is also displayed in the same coordinate system. Figure 13 shows the accumulated background and one of the IMOs, with the associated trajectories (represented as the red and green string of arrows); (a) shows only the accumulated background and the egomotion; (b) and (c) also show the accumulated IMO and its trajectory; (d) shows the IMO's estimated trajectory, from above. The car's egomotion averages to 65 km/h on this sequence, whereas the IMO's mo-
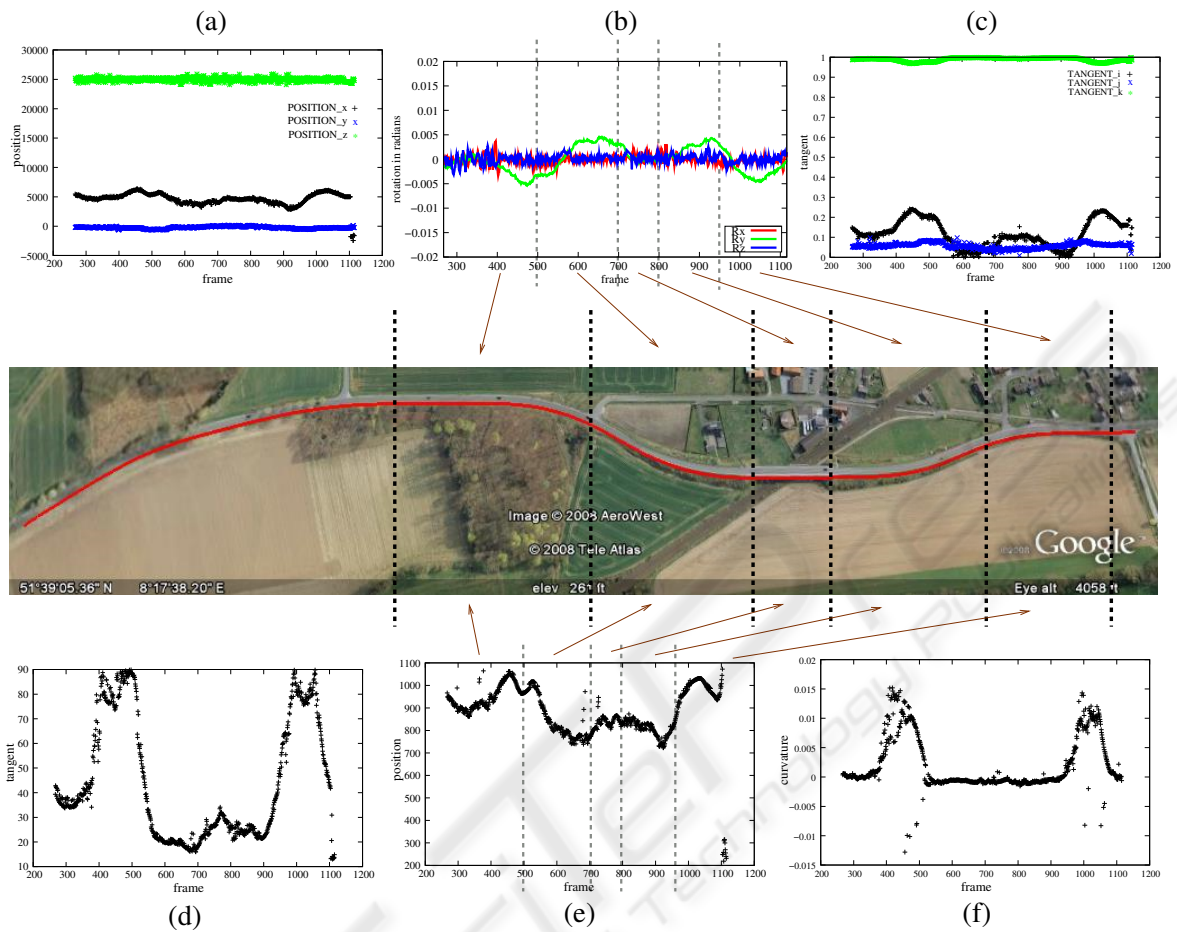
Figure 10: Motion of the car from different sources (2D-3D lane parameters and ego motion estimation). (a)-(c) (x,y,z) coordinates and components of the tangent vector of the 3D lane for $z = 25 meters$ on the whole reference sequence. (b) Ego motion estimation for rotation by using combination of SIFT and primitives. (d-f) 2D lane parameters for $y = 500 pixels$ on the whole reference sequence. (d) slope (e) x coordinate (f) curvature.

Figure 11: IMOs processing. (a) IMO list and BBs (b) left image segment, (c) right image segment, (d) extracted primitives and stereo correspondences.
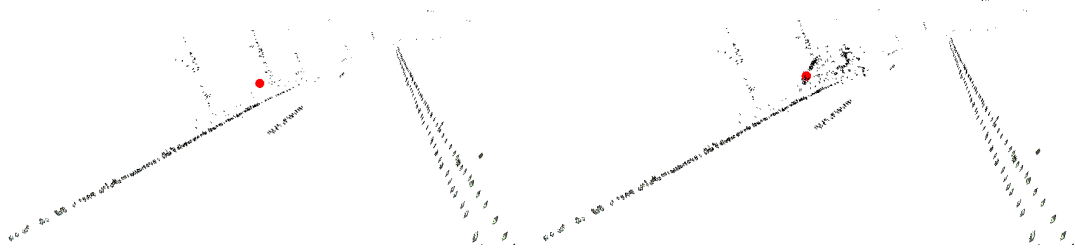
503

Figure 12: Segmentation and reconstruction of an IMO: left shows the accumulated map of the road, where IMOs are segmented out; right shows the accumulated representation of the IMO, on the same coordinate system.
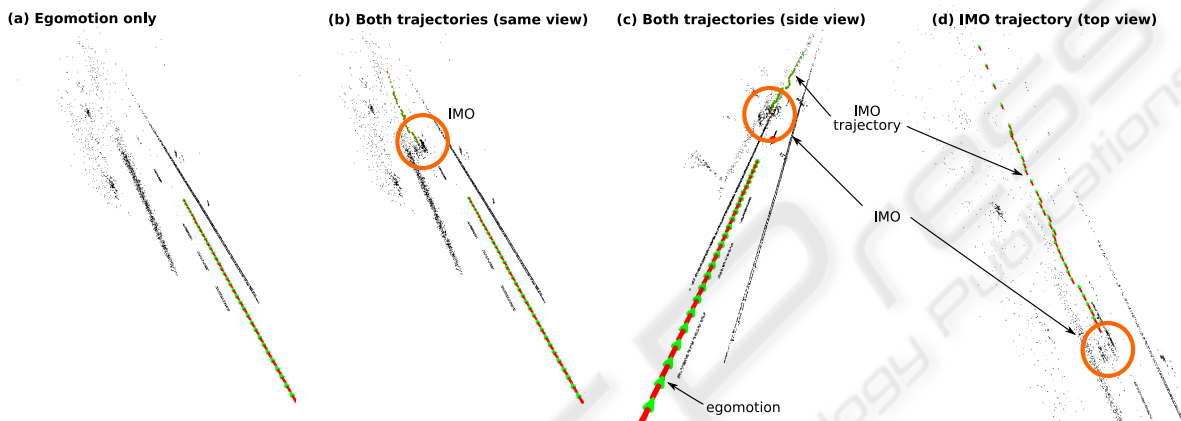

Figure 13: One IMO in a semantic map and its trajectory. (a) shows only the accumulated background and the car's egomotion; (b) and (c) show in addition the accumulated IMO and its trajectory, under different viewpoints; and (d) shows the IMO's trajectory viewed from above.

tion averages at 80 km/h. In those results one see that the IMOs can be segmented and tracked. The use of dense method for detection and tracking provides robustness, and the use of symbolic features for accumulation provides consistency to the system.

8 CONCLUSIONS

We have discussed four algorithms as sub-parts of a road interpretation, based on visually extracted data, provided by an early cognitive vision system. After increasing the reliability of 2D and 3D information by using temporal and spatial disambiguation mechanisms, lane markers of the road are detected and the ego-motion of the car is estimated. While having the ego-motion in hand, large scale maps of the road and independently moving objects on the road are created. The potential of the presented algorithms has been demonstrated on a stereo image sequence, collected from a structured road. We have shown that the resultant large scale map, as well as the estimated motion and the detected lane parameters are consistent with the satellite image. Note that, all sub-tasks use only

visually extracted information and all parts of this information are provided by the same representation.

REFERENCES

Başeski, E., Pugeault, N., Kalkan, S., Kraft, D., Wörgötter, F., and Krüger, N. (2007). A Scene Representation Based on Multi-Modal 2D and 3D Features. *ICCV 2007 Workshop on 3D Representation for Recognition 3dRR-07*.

Bertozzi, M. and Broggi, A. (1998). GOLD: A Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection. *Image Processing, IEEE Transactions on*, 7(1):62–81.

Bertozzi, M., Broggi, A., and Fascioli, A. (2000). Vision-based Intelligent Vehicles: State of the Art and

Perspectives. *Robotics and Autonomous Systems*, 32(1):1–16.

Faugeras, O. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press.

Fischler, R. and Bolles, M. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):619–638.

Gordon, N., Salmond, D., and Smith, A. (1993). Novel Approach to nonlinear/non–Gaussian Bayesian State Estimation. In *IEE Proceedings-F*, volume 140, pages 107–113.

Hermann, S. and Klette, R. (2008). A Study on Parameterization and Preprocessing for Semi-Global Matching. Technical Report Number 221, Computer-Science Department of The University of Auckland, CITR.

Jensen, L. B. W., Kjær-Nielsen, A., Alonso, J. D., Ros, E., and Krüger, N. (2008). A Hybrid FPGA/Coarse Parallel Processing Architecture for Multi-modal Visual Feature Descriptors. *ReConFig'08*.

Klette, R. (2008). Stereo-Vision-Support for Intelligent Vehicles - The Need for Quantified Evidence. Technical Report Number 226, Computer-Science Department of The University of Auckland, CITR.

Krüger, N., Lappe, M., and Wörgötter, F. (2004). Biologically Motivated Multi-modal Processing of Visual Primitives. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*, 1(5):417–428.

Leibe, B., Cornelis, N., Cornelis, K., and Van Gool, L. (2007). Dynamic 3D Scene Analysis from a Moving Vehicle. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8.

Lemaire, T., Berger, C., Jung, I.-K., and Lacroix, S. (2007). Vision–Based SLAM: Stereo and Monocular Approaches. *International Journal of Computer Vision*, 74(3):343–364.

Lowe, D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 2(60):91–110.

McCall, J. and Trivedi, M. (2004). An Integrated, Robust Approach to Lane Marking Detection and Lane Tracking. pages 533–537.

Pauwels, K. (2008). *Computational Modeling of Visual Attention:Neuronal Response Modulation in the Thalamocortical Complex and Saliency-based Detection of Independent Motion*. PhD thesis, K.U.Leuven.

Pauwels, K. and Van Hulle, M. M. (2008). Optic Flow from Unstable Sequences through Local Velocity Constancy Maximization. *Image and Vision Computing*, page in press.

Piegl, L. and Tiller, W. (1995). *The NURBS Book*. Springer-Verlag, London, UK.

Pugeault, N., Wörgötter, F., and Krüger, N. (2006). Multi-modal Scene Reconstruction Using Perceptual Grouping Constraints. In *Proc. IEEE Workshop on Perceptual Organization in Computer Vision (in conjunction with CVPR'06)*.

Pugeault, N., Wörgötter, F., and Krüger, N. (2008). Accumulated Visual Representation for Cognitive Vision. *In Proceedings of the British Machine Vision Conference (BMVC)*.

Rosenhahn, B., Krüger, N., Rabsch, T., and Sommer, G. (2001). Automatic tracking with a novel pose estimation algorithm. *Robot Vision 2001*.

Wang, Y., Teoh, E., and Shen, D. (2004). Lane Detection and Tracking Using B-Snake. 22(4):269–280.