# Semantic Reasoning for Scene Interpretation

Lars B.W. Jensen†, Emre Baseski†, Sinan Kalkan‡, Nicolas Pugeault±, Florentin Wörgötter‡ and Norbert Krüger†

†University of Southern Denmark
Odense, Denmark
{lbwj, emre, norbert}@mmmi.sdu.dk

‡University of Göttingen
Göttingen, Germany
{sinan, worgott}@bccn-goettingen.de

± University of Edinburgh
Edinburgh, United Kingdom
npugeaul@ed.ac.uk

**Abstract.** In this paper, we propose a hierarchical architecture for representing scenes, covering 2D and 3D aspects of visual scenes as well as the semantic relations between the different aspects. We argue that labeled graphs are a suitable representational framework for this representation and demonstrate its potential by two applications. As a first application, we localize lane structures by the semantic descriptors and their relations in a Bayesian framework. As the second application, which is in the context of vision based grasping, we show how the semantic relations can be associated to actions that allow for grasping without using any object knowledge.

## 1 Introduction

In this work, we represent scenes with a hierarchy of visual information. The input consists of stereo images (or sequences of them) that become processed at different levels. Information of increasing semantic richness becomes processed at the different levels, covering multiple aspects of a scene such as 2D and 3D information as well as geometric and appearance based information. Furthermore, the spatial extent of the processed entities increases in the higher levels of the hierarchy.

We make use of rich local symbolic descriptors, describing edge-like structures and homogeneous structures, as well as groups (contours and areas) formed by them. Furthermore, rich semantic relations between these descriptors and the groups are defined. The descriptors describe local information in terms of multiple visual modalities (2D and 3D position and orientation, colour as well as contrast transition). Moreover, there is a set of semantic relations defined between them such as the Euclidean distance in 2D and 3D as well as parallelism, co-planarity and co-colority (i.e., sharing similar colour structure).

Scenes become represented as a set of labeled graphs, whose nodes are labeled by properties of local descriptors, groups and areas thereof and edges between the nodes represent the semantic relations between the nodes in the graphs.

Idealized graphs can be defined or learned from scene structures such as road lanes and can be efficiently matched with the extracted scene graphs by making use of the rich semantics.

From a cognitive point of view, it is important to have a representation that allows for an efficient storage of information as well as for reasoning processes on visual scenes. From a storage point of view, it is not convenient to memorize information on a very low and local level since it would require a large amount of memory. Also it would be much more difficult for learning processes to make use of relevant semantics. As a consequence, the very condensed graph representation is much more suitable for memorizing objects.

We present two applications of our hierarchical framework: As a first application, we show how a street structure can be characterized by both its appearance and relations between its sub-components. Here, the matching process is governed by Bayesian reasoning based on local descriptors and semantic relations between them, which are controlled by prior probabilities. Moreover, this Bayesian reasoning process makes explicit the relative importance of the different cues and relations opening the way for the learning of sparse graph structures. In terms of semantic reasoning, we can show that, by means of the semantic relations, it is possible to mediate between textual descriptions of scene structures (e.g., the lanes) and visual detection as examplified. Such graphs can be idealized (or, generalized) either through learning or can be provided as world knowledge, and be used for matching (see section 4.1).

The second application is based on [1], and illustrates how the approach presented herein embeds in a robotic scenario. In this scenario, groups of visual features fulfilling certain semantic relations can be associated to grasping actions, allowing for the grasping of objects without using any model knowledge.

The use of hierarchical representations, mostly graphs, is commonplace for scene representation. For example, *scene graphs* and *spatial relationship graphs* are heavily used in Computer Graphics for representing 3D world and scenes [2]; such graphs are designed mostly for rendering purposes, and they are not sufficient for covering the 2D properties of scenes. *Relative Neighborhood Graph*s, introduced by [3], are used in Computer Vision studies for representation of structured entities [4]. A similar graphical structure called *Region Adjacency Graph* is used for region-based representation of objects or scenes [5, 6]. There exist a variety of similar graphical representations and we refer the interested reader to [7].

Our contribution in this paper is the introduction of a hierarchical vision system that allows for semantic reasoning based on rich descriptors and their relations. This vision system covers not only the appearance aspects but also the geometrical properties of the scene, which allows for doing reasoning in both 2D and 3D world. In particular, it allows for the step-wise translation of a textual description of an object to a visual representation that can be used for localizing a certain structure in a visual scene.

The paper is structured as follows: In section 2, the visual scene representation is introduced. In section 3, we describe the embedding of the visual representation in graphs. We then describe the two applications in section 4. We introduce the algorithm for the detection of a lane structure in section 4. Another application in the context of vision based grasping is described in section 4.2. In section 5, we discuss the potential of this approach in terms of a cognitive system architecture.

## 2 Hierarchical Architecture

We represent scenes with a three-level architecture of visual entities (see figure 1) of increasing richness and semantic. In the following subsections, we introduce the different levels of this hierarchical representation in order of increasing complexity, starting from the lowest level.

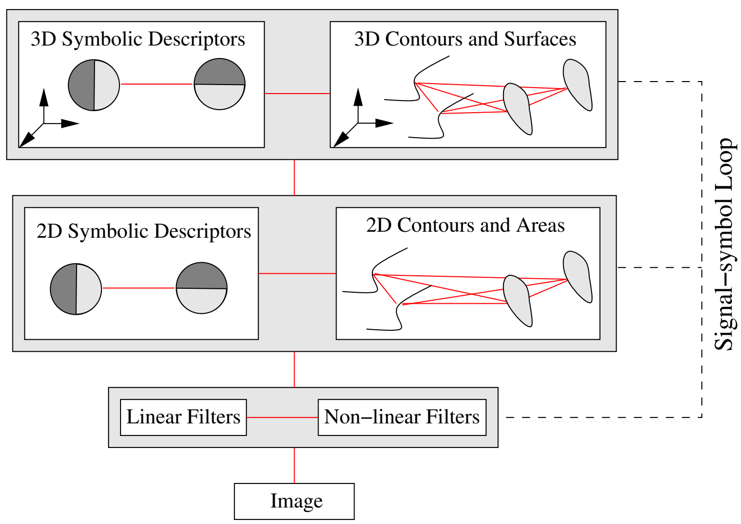


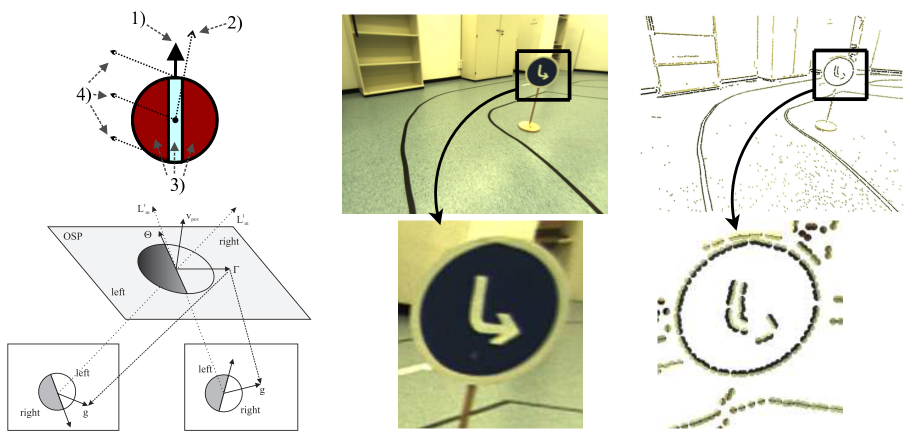

**Fig. 1.** An overview of the hierarchical architecture introduced in this paper. The visual entities denote the nodes of the graphical representation, and the red edges, which correspond to perceptual grouping and correspondence relations, are the links between the nodes. Higher levels in the hierarchy correspond to more symbolic, more spacious and more descriptive visual entities. See the text for more details and figure 6 for examples of the different levels of the hierarchical architecture.

### 2.1 Linear and non-linear filtering

At the first level, we apply a combination of linear and non-linear filtering operations to extract pixel-wise signal information in terms of local magnitude, orientation, phase [8] as well as optical flow [9] — for details see [10, 11].

## 2.2 Symbolic Representation in 2D

The transition to a local symbolic description is done at the second level (the "Symbolic Representation in 2D" layer in figure 6) where local image patches are described by the so-called *multi-modal primitives* [12]. The primitives provide a condensed semantic description of the local (spatial-temporal) signal in terms of image orientation, phase, colour and optic flow. The difference to the first level is that the information is sparsified, highly condensed and associated to discrete positions with sub-pixel accuracy. Figure 2 shows extracted 2D primitives (denoted as $\pi$) for an example scene.




(a) (b) (c)

**Fig. 2. (a)** Representation and attributes of a 2D primitive where (1) orientation of the primitive, (2) the phase, (3) the color and (4) the optic flow and reconstruction of a 3D primitive. **(b)** A sample scene and a closer view for the region of interest. **(c)** Extracted 2D primitives for the example scene in **(b)**.

At this level, the information is sparsely coded such that interaction processes between visual events can be modeled more efficiently than at the pixel level (for a detailed description of these interaction processes see, e.g., [13]). Already at this level, semantic relations between local 2D primitives can be defined. Besides the 2D distance, primitives allow collinearity and co-colourity relations to be defined between them: Two primitives are collinear if they are part of the same line (figure 3(a) and 4(f)). Two primitives, on the other hand, are co-colour if the colours of their *sides* that face each other are similar (figures 3(c) and 4(e)). See [14] for more information about the definition of these relations.

The 2D descriptors naturally organize themselves along contours and the semantic description is highly correlated along such a contour (e.g., 2D orientation varies smoothly and in general colour, phase and optic flow are similar for the primitives on the contour). Hence, it is natural to condense the information of

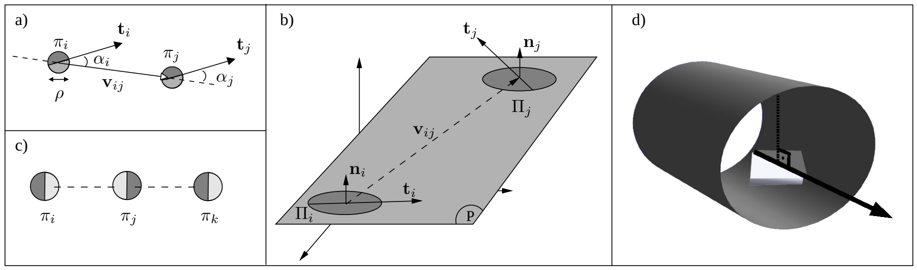



**Fig. 3.** Illustration of the perceptual relations between primitives. **(a)** Collinearity of two 2D primitives. **(b)** Co–planarity of two 3D primitives $\boldsymbol{\Pi}_i$ and $\boldsymbol{\Pi}_j$. **(c)** Co–colority of three 2D primitives $\pi_i, \pi_j$ and $\pi_k$. In this example, $\pi_i$ and $\pi_j$ are cocolor, so are $\pi_i$ and $\pi_k$; however, $\pi_j$ and $\pi_k$ are not cocolor. **(d)** Normal distance between $\boldsymbol{\Pi}_i$ and $\boldsymbol{\Pi}_j$ is 0 if $\boldsymbol{\Pi}_j$ is outside the cylindrical volume surrounding $\boldsymbol{\Pi}_i$ and defined otherwise as the distance between $\boldsymbol{\Pi}_j$ and the line created from the location of $\boldsymbol{\Pi}_i$ which goes in the direction of $\boldsymbol{\Pi}_i$s orientation vector.

the primitives organized along a contour in the form of a more abstract parameterization in terms of unified appearance based descriptors as well as a NURBS (Non-Uniform Rational B-Splines [15]) representation of the geometry of the contours (see figure 5). By this, we reduce the number of bits used to represent a scene further as well as the number of second order relations of visual events. The latter point is in particular relevant, when we want to code objects with these relations.

### 2.3 Symbolic Representation in 3D

Using the corresponding 2D primitives in the left and right image, 3D primitives can be reconstructed (denoted by $\boldsymbol{\Pi}_j$). At the third level, the reconstructed 3D primitives inherit the appearance based properties of the 2D primitives (phase and colour) and extend the 2D position and 2D orientation to 3D (see figure 2). Moreover, the semantic relations between 2D primitives can be extended to the 3D primitives and also further enriched by particular 3D relations such as co-planarity or 3D properties such as in-ground-plane (see figures 3 and 4). Co-planarity refers to the *being-on-the-same-plane* relation between two 3D primitives or 3D contours (figures 3(b) and 4(d)). See [14] for more information about the definition of co-planarity. In-ground-plane relation, on the other hand, corresponds to all 3D entities that are in the ground plane (figure 4(c)). The 2D contour representation becomes also extended to 3D contours by connecting 3D primitives that are linked together. NURBS are fitted to the 3D contours as in 2D to obtain a global mathematical description of the 3D contours. In addition, the NURBS parametrization can be used to increase the precision of the local feature extraction process (see figure 5).

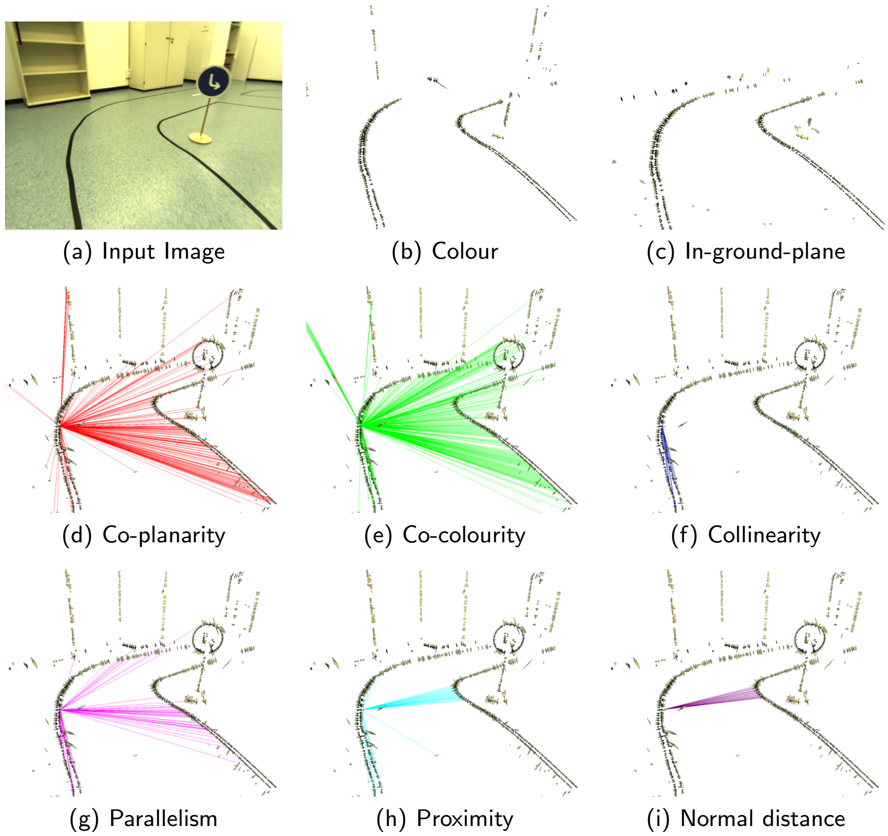



**Fig. 4.** A set of 2D and 3D relations for the visual entities extracted from an example scene whose left view is provided in (a). (b) Primitives which are black. (c) 3D primitives which satisfy the "ground-plane" relation. (d-g) Connects the 3D primitives that are respectively co-planar, co-colour, collinear and parallel to a selected 3D primitive. (h) Connects the 3D primitives that are of a given 3D distance to a selected 3D primitive. (i) Connects the 3D primitives whose normal distance to a selected 3D primitive equals a given value.

Note that this process is not a pure bottom-up process, as it involves corrective feedback mechanisms at various levels. These are described in more detail in, e.g., [13, 16].

## 3 Semantic Graphs

The hierarchy of representations discussed above provides us with a number of 2D and 3D local entities that are linked to more global entities. These entities are semantically rich as such, and in addition there exist semantic relations between them. Because of this linkage, we suggest that labeled graphs are the suitable

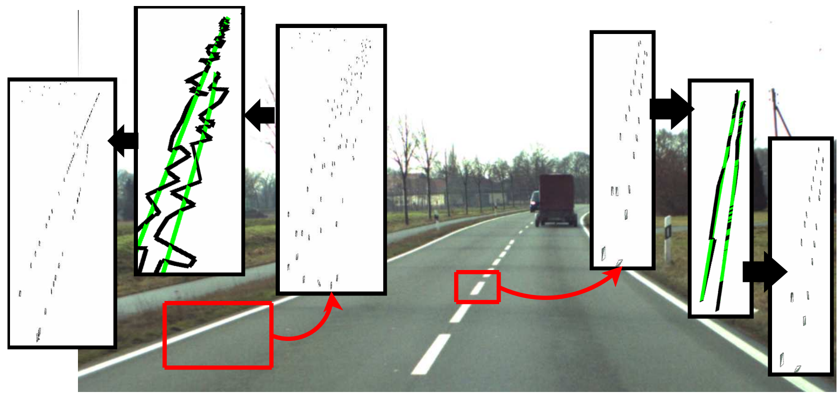


**Fig. 5.** Position and orientation correction of 3D primitives by using NURBS. After fitting NURBS (represented as green lines) to groups of primitives (represented as black lines), position and orientation of each primitive is recalculated. The procedure is shown on a good reconstruction (middle road marker) as well as a bad one (left lane marker).

representational framework for representing scenes. In these graphs, the nodes represent different visual entities such as primitives, contours and areas with their first order properties while the links represent the semantic relations. Note that actually we have a set of labeled graphs, which are linked to each other and with this linkage, they cover the 2D and 3D aspects of a scene (see figure 6) since each relation naturally defines a sub-graph, covering a structure in a scene.

In processing of information across the different levels, the semantic richness of information increases from level to level. However, it is important to point out that with this increase of semantical richness, also the likelihood of errors in the processing increases due to loss of valuable information or introduction of noise through thresholding. In addition, the uncertainty of visual information, in particular in the 3D domain, might also make any reasoning uncertain. Hence, we intend to be able to use the extracted information *on all levels* according to the current task and uncertainties of information at the different levels. In addition, spatial–temporal processes are defined that increase the stability and the certainty of information by spatial-temporal predictions [13]. The proposed hierarchy allows for processes that transfer information from the symbolic level to the signal level to recover weak information in so-called signal-symbol loops (see [16]). Such loops are essentially feedback mechanisms that carry the results of symbolic processing to the signal level.

## 4 Applications

In this section, we give two applications of the semantic reasoning process. First, we show how a lane structure can be described by the semantic descriptors and

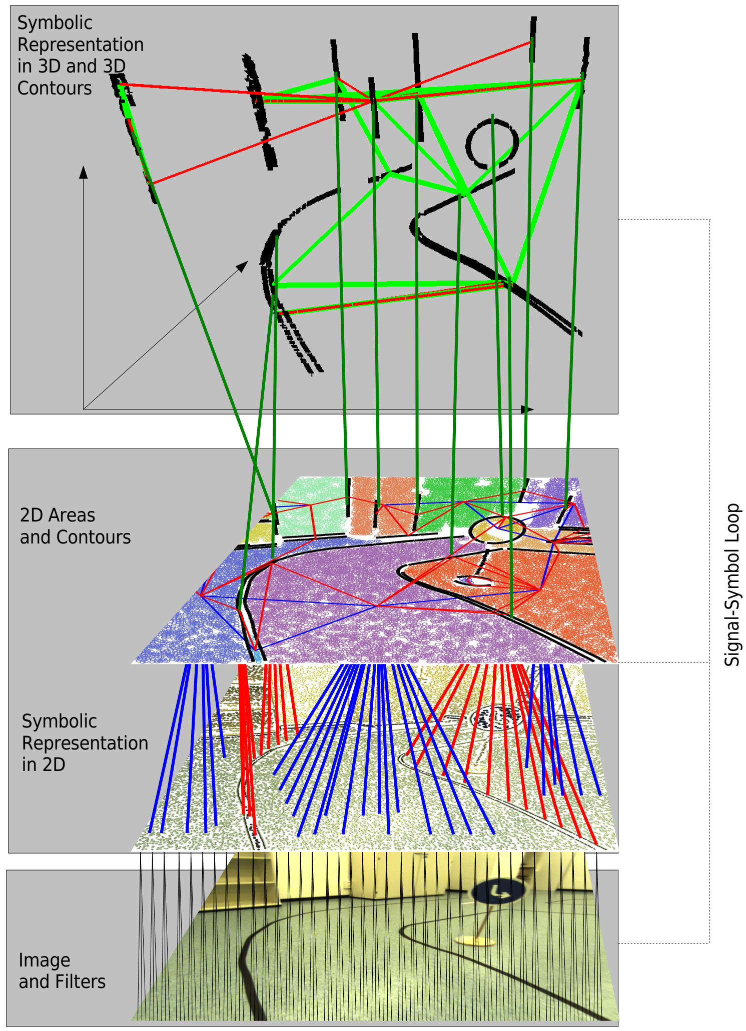



**Fig. 6.** A multi-level graph structure. For clarity, only a subset of the links is drawn, and the links corresponding to different relations such as parallelism and co-colority between 2D or 3D entities are skipped. “Image and Filters” (IF) layer is the input image which contains pixels as the nodes of the graph. “Symbolic Representation in 2D” (SR-2D) layer contains the 2D primitives. The links between the IF layer and the SR-2D layer correspond to ”part-of” relations between pixels and primitives. “2D Contours and Areas” (CA-2D) layer contains image areas (each area is drawn in a different color) and 2D contours (in black). The neighborhood relations between two areas and between an area and a contour are drawn respectively in blue and red. The links between the SR-2D layer and the CA-2D layer correspond to ”part-of” relations between primitives, and areas and contours. The “Symbolic Representation in 3D and 3D Contours” (SRC-3D) layer includes 3D contours in black (the 3D surfaces are skipped for clarity), and the links in red and light green between the 3D contours respectively denote coplanarity and cocolority relations between the contours. The links between the CA-2D layer and the SRC-3D layer are ”projection” relations between the 2D and 3D contours.

their relations in a Bayesian framework (section 4.1). Then we describe another application in a robotic context (section 4.2).

### 4.1 Lane finding using Bayesian Reasoning

A lane in our lab environment (see figure 4(a)) can be characterized by the colour and the width of the lane marker, which is known also to be in the ground plane, as well as by its distance to the other lane marker. As a textual description of the lane one could state:

> A lane consists of two lane markers with distance $d_{far}$ which are both in the ground plane. A lane marker has a width $d_{near}$ and has the colour 'black'.

An idealized representation of this textual description in a graph is shown in figure 7. The representation introduced in the last two sections allows for directly applying the terms used in the textual description. Colour and 'being in ground plane' are first order attributes of primitives and groups while the term 'distance' corresponds to the relation 'normal distance' (figure 3). Hence, the textual description can be easily translated in our visual representations. However, there are two problems we have to face: First, a lane is not described by one property, or relation, but by a number of properties. Therefore, these different cues need to be combined. Second, scene interpretation processes have to face uncertainties in the feature extraction process. Reasons for the uncertainties are, for example, noise in the recording process, limited resolution as well as the correspondence problem in the stereo reconstruction.

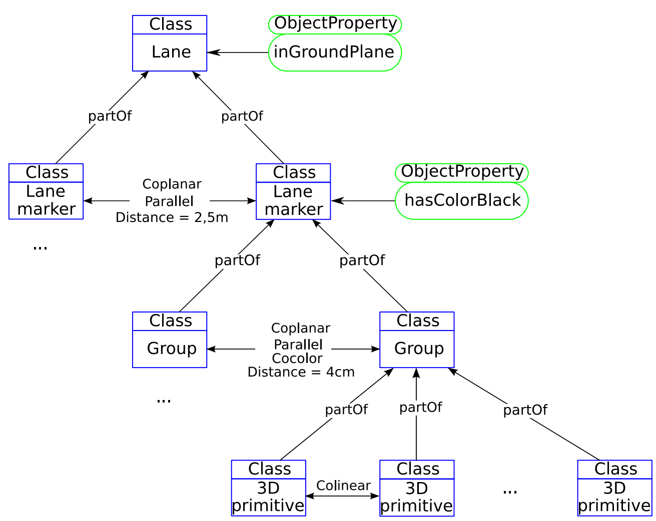



**Fig. 7.** A graph showing an idealized representation of the lane in our lab environment.

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

– making explicit statements about the relevance of properties for a certain object,
– introduction of learning in terms of prior and conditional probabilities, and
– assessing the relative importance of each type of relation for the detection of a given object, using the conditional probabilities.

Bayes formula (e.g., see [17]) enables to infer the probability of an unknown event conditioned to other observable events and to prior likelihoods. Let $P(e_i^{\Pi})$ be the prior probability of the occurrence of an event $e_i^{\Pi}$ (e.g., the probability that any primitive lies in the ground plane). Then, $P(e_i^{\Pi}|\Pi \in \mathcal{O})$ is the conditional probability of the visual event $e_i$ given an object $\mathcal{O}$.

Our aim is to compute the likelihood of a primitive $\Pi$ being part of an object $\mathcal{O}$ given a number of visual events relating to the primitive:

$$P(\Pi \in \mathcal{O}|e_1^{\Pi}, \ldots, e_n^{\Pi}). \tag{1}$$

According to Bayes formula, equation 1 can be expanded to:

$$\frac{P(e_1^{\Pi}, \ldots, e_n^{\Pi}|\Pi \in \mathcal{O})P(\Pi \in \mathcal{O})}{P(e_1^{\Pi}, \ldots, e_n^{\Pi}|\Pi \in \mathcal{O})P(\Pi \in \mathcal{O}) + P(e_1^{\Pi}, \ldots, e_n^{\Pi}|\Pi \neg \in \mathcal{O})P(\Pi \neg \in \mathcal{O})}. \tag{2}$$

In this work we assume independence between $e_1^{\Pi}, \ldots, e_n^{\Pi}$ (we intend to investigate to what degree this assumption holds in a future work). If $e_1^{\Pi}, \ldots, e_n^{\Pi}$ are independent then $P(e_1^{\Pi}, \ldots, e_n^{\Pi}|\Pi \in \mathcal{O})$ can be written as:

$$P(e_1^{\Pi}, \ldots, e_n^{\Pi}|\Pi \in \mathcal{O}) = P(e_1^{\Pi}|\Pi \in \mathcal{O}) \cdot \ldots \cdot P(e_n^{\Pi})|\Pi \in \mathcal{O}), \tag{3}$$

and

$$P(e_1^{\Pi}, \ldots, e_n^{\Pi}|\Pi \neg \in \mathcal{O}) = P(e_1^{\Pi}|\Pi \neg \in \mathcal{O}) \cdot \ldots \cdot P(e_n^{\Pi}|\Pi \neg \in \mathcal{O}), \tag{4}$$

and the formula (2) becomes rather easy.

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 relevant relations for these selections. The results are shown in table 1. The numbers reveal that 'being in ground plane' and 'near normal distance' are the strongest relations as they show the largest difference in probability between the conditions 'in lane' and 'not in lane'.

Figure 8 shows the results of using the Bayesian framework with the computed prior probabilities in two different scenarios: our indoor lab environment and an outdoor scene. The same prior probabilities were used in both scenarios, but for the outdoor scene, the values and thresholds of the relations underlying the probabilities had to be changed to fit the color and dimensions of a real lane.

**Table 1.** Prior probabilities.

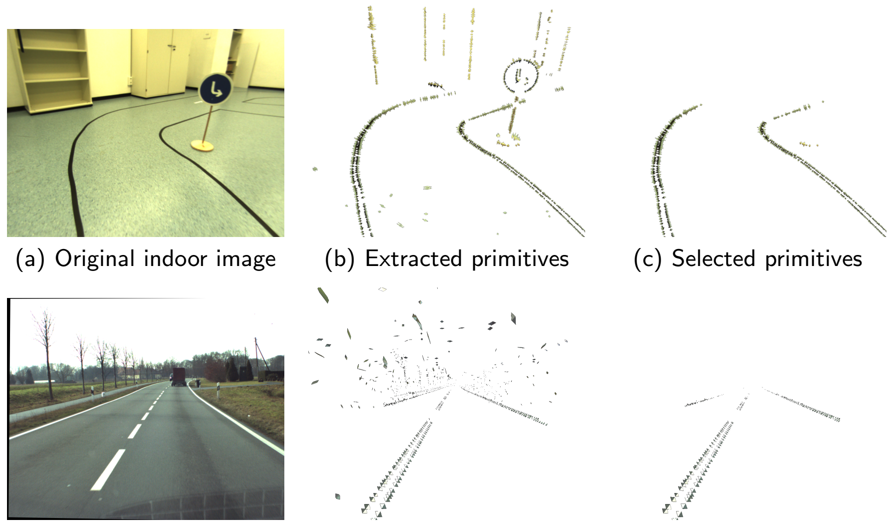
| Type | Probability |
|---|---|
| $P(\Pi\ in\ lane)$ | 0.44792 |
| $P(\Pi\ not\ in\ lane)$ | 0.55208 |
| $P(\Pi\ being\ black)$ | 0.70058 |
| $P(\Pi\ being\ black \mid \Pi\ in\ lane)$ | 0.97959 |
| $P(\Pi\ being\ black \mid \Pi\ not\ in\ lane)$ | 0.47391 |
| $P(\Pi\ in\ ground\ plane)$ | 0.49925 |
| $P(\Pi\ in\ ground\ plane \mid \Pi\ in\ lane)$ | 0.95960 |
| $P(\Pi\ in\ ground\ plane \mid \Pi\ not\ in\ lane)$ | 0.12543 |
| $P(\Pi\ has\ normal\ distance\ d_{far})$ | 0.35943 |
| $P(\Pi\ has\ normal\ distance\ d_{far} \mid \Pi\ in\ lane)$ | 0.66433 |
| $P(\Pi\ has\ normal\ distance\ d_{far} \mid \Pi\ not\ in\ lane)$ | 0.11131 |
| $P(\Pi\ has\ normal\ distance\ d_{near})$ | 0.41015 |
| $P(\Pi\ has\ normal\ distance\ d_{near} \mid \Pi\ in\ lane)$ | 0.86170 |
| $P(\Pi\ has\ normal\ distance\ d_{near} \mid \Pi\ not\ in\ lane)$ | 0.04377 |

### 4.2 Associating Actions to Co-planar Groups

To underline the embedding and strength of our approach of utilizing semantic relations between visual events in the hierarchical representation described in section 2, we briefly present new results on an application that has been described in more detail in [1]. In this application, relations between primitives (or groups) become associated to actions. In figure 9 (left bottom), a grasping hypothesis connected to a co-planar pair of primitives is shown. Hence, the co-planarity graph shown in figure 9 (right), corresponding to the white butter dish, can be associated to grasping hypotheses (as indicated in the middle of the figure). In [18], we could show that by such a simple mechanism, objects in rather complex scenes can be grasped with a high success rate. In figure 10 (left), a scene with a number of objects is shown. Using the grasping reflex described in 9, it was possible to clean the scene (after approximately 30 grasping attempts) except one object for which the system's embodiment precluded grasping (i.e., the two finger grasper attachment of the robot could not grasp the round can in any way).

## 5 Discussion

In this work, we introduced a hierarchical representation of semantically rich descriptors and their relations, and argued that labeled graphs are a suitable framework for scene representation, enabling cue merging and action association. Within this representation, Bayesian reasoning has been applied for efficient cue-merging, allowing for relating textual descriptions to extracted visual information. We also outlined that in such a framework feedback mechanisms at different levels can be used to disambiguate the information, in particular through feedback between the symbolic and signal level.



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

(d) Original outdoor image (e) Extracted primitives (f) Selected primitives

**Fig. 8.** Extracting the lane in two scenarios: (a-c) showing our indoor lab environment and (d-f) showing an outdoor scenario.

In our current work, we are aiming at the development of efficient matching strategies that realize the full potential of our representations. In particular, we are interested in structures that cannot be completely defined by their appearance only (as for example in the case of street signs) but by the relations of sub-structures to each other (as, for example, in case of the task of distinguishing different kinds of road structures such a motorways, crossings, motorway exits but also in other more general object categorization tasks).


## 6 Acknowledgements

This work has been supported by the European Commission - FP6 Project DRIVSCO (IST-016276-2).


## References


1. Aarno, D., Sommerfeld, J., Kragic, D., Pugeault, N., Kalkan, S., Wörgötter, F., Kraft, D., Krüger, N.: Early reactive grasping with second order 3D feature relations. In Lee, S., Suh, I.H., Kim, M.S., eds.: Recent Progress in Robotics; ViableRobotic Service to Human, selected papers from ICAR'07. Springer-Verlag Lecture Notes in Control and Information Sciences (LNCIS) (2007)
2. Echtler, F., Huber, M., Pustka, D., Keitler, P., Klinker, G.: Splitting the scene graph – using spatial relationship graphs instead of scene graphs in augmented reality. In: GRAPP'08: Int. Conference on Computer Graphics Theory and Applications. (2008)

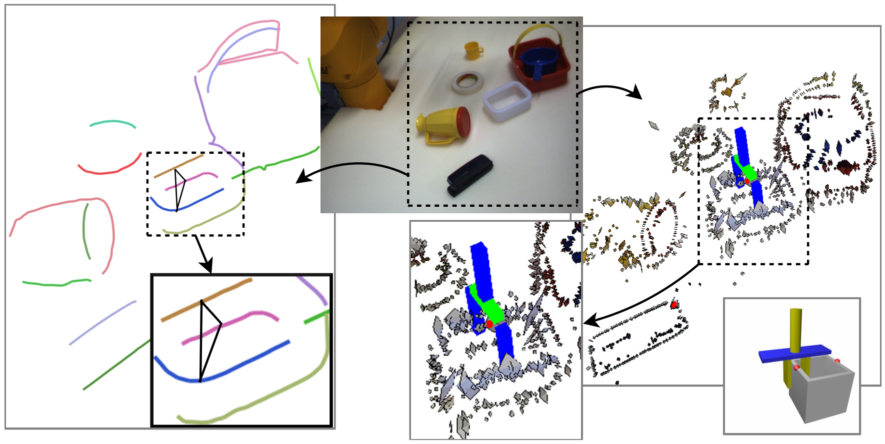
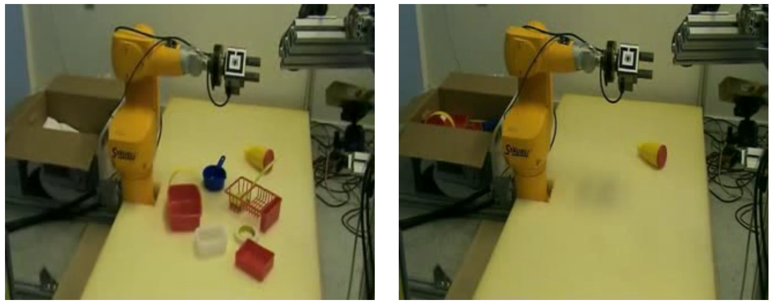


**Fig. 9.** The 2D contours extracted from the example view on the top-middle are drawn in different colors on the left. The coplanarity graph of the white cup is also shown in black on the left, and this graph suggests a grasp of the type shown in the lower right (the red spheres represent two coplanar representative primitives out of the two contours). The resulting grasp is shown on the left and in the bottom-middle image.


3. Jaromczyk, J.W., Toussaint, G.T.: Relative neighborhood graphs and their relatives. Proceedings of the IEEE **80**(9) (Sep 1992) 1502–1517
4. Mucke, E.P.: Shapes and implementations in three-dimensional geometry. Technical report, University of Illinois at Urbana-Champaign, Champaign, IL, USA (1993)
5. Korting, T.S., Fonseca, L.M.G., Dutra, L.V., da Silva, F.C.: Image re-segmentation – a new approach applied to urban imagery. In: VISAPP'08: Int. Conference on Computer Vision Theory and Applications. (2008)
6. Tremeau, A., Colantoni, P.: Regions adjacency graph applied to color image segmentation. IEEE Transactions on Image Processing **9**(4) (2000) 735–744
7. Hancock, E.R., Wilson, R.C.: Graph-based methods for vision: A yorkist manifesto. In: Proc. of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, London, UK, Springer-Verlag (2002) 31–46
8. Kovesi, P.: Image features from phase congruency. Videre: Journal of Computer Vision Research **1**(3) (1999) 1–26
9. Nagel, H.H.: On the estimation of optic flow: Relations between different approaches and some new results. Artificial Intelligence **33** (1987) 299–324
10. Sabatini, S.P., Gastaldi, G., Solari, F., Diaz, J., Ros, E., Pauwels, K., Hulle, K.M.M.V., Pugeault, N., Krüger, N.: Compact and accurate early vision processing in the harmonic space. International Conference on Computer Vision Theory and Applications (VISAPP) (2007)
11. Felsberg, M., Sommer, G.: The monogenic signal. IEEE Transactions on Signal Processing **49**(12) (December 2001) 3136–3144
12. Krüger, N., Lappe, M., Wörgötter, F.: Biologically motivated multi-modal processing of visual primitives. Interdisciplinary Journal of Artificial Intelligence & the Simulation of Behaviour, AISB Journal **1**(5) (2004) 417–427



(a) (b)

**Fig. 10.** Co-planar pairs of contours predict groups. (a) The four different elementary grasping actions defined based on a pair of co-planar groups. (b) Robot scene before the grasping procedure has been applied. (c) Scene after all graspable Objects have been removed by the system.


13. Pugeault, N.: Early Cognitive Vision: Feedback Mechanisms for the Disambiguation of Early Visual Representation. PhD thesis, Informatics Institute, University of Göttingen (2008)
14. Kalkan, S., Pugeault, N., Krüger, N.: Perceptual operations and relations between 2d or 3d visual entities. Technical Report 2007-3, Technical report of the Robotics Group, Maersk Institute, University of Southern Denmark (2007)
15. Piegl, L., Tiller, W.: The NURBS book (2nd ed.). Springer-Verlag New York, Inc., New York, NY, USA (1997)
16. Kalkan, S., Yan, S., Krüger, V., Wörgötter, F., Krüger, N.: A signal-symbol loop mechanism for enhanced edge extraction. In: VISAPP'08: Int. Conference on Computer Vision Theory and Applications. (2008)
17. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, Inc. (1988)
18. Popović, M.: An early grasping reflex in a cognitive robot vision system. Master's thesis, University of Southern Denmark (2008)