Context-Aware Object Priors

Alexander Bachmann and Michael Balthasar
Department for Measurement and Control
University of Karlsruhe (TH)
76 131 Karlsruhe, Germany
bachmann@mrt.uka.de

***Abstract*— In this contribution, we present a probabilistic model for video-based object recognition that jointly considers object and context information. The model can be interpreted as an extension of the Bayesian framework, which forms the basis for most probabilistic object recognition methods. We consider the environment in the vicinity of the observer as a profitable source of context information and propose a prior that is capable of imposing such scene information into the object recognition process. After introducing the context descriptor used in this work, a simple model is presented to express different scene categories within the road-traffic domain. We show how context information can be incorporated into our existing object recognition scheme by evaluating the dependencies between a specific scene category and selected object properties. The proposed descriptor has been trained and the model has been evaluated on real image data.**

I. Introduction

Autonomous navigation in open and dynamic environments implies the comprehensive perception and understanding of the environment in the vicinity of the vehicle. Especially for applications in the road traffic domain the robust and reliable detection of close-by traffic participants is of major interest. In this context, vision sensors provide a rich and versatile source of information [9], [7]. Based on the finding that the human visual system not only makes use of local, object-inherent information for object detection and recognition but also considers the global scene context, we propose an object recognition method that jointly considers local and global image information. Studies on human scene perception have shown that the structural arrangement of many real-world scenes has a stronger impact to the overall understanding of an environment and the objects contained within than the identification and analysis of the objects itself. An empirical proof of this can be found in [2]. For the development of reliable and comprehensive computer vision systems – aiming to achieve the level of human performance – we believe that it is essential to broaden ones attention beyond the object itself considering the entire image for the global recognition task.

In previous studies the contextual information has been expressed by e.g. defining a collection of objects and a set of rules about the world in which the system is expected to operate [3], [11]. In [10] a statistical approach is presented that learns the joint distribution of a number of objects within the scene based on a set of local measurements in the image. The coarse geometric property of a scene is estimated in [4] by learning appearance-based models of geometric classes. Another possibility to represent context information is to consider it as a single entity that can be recognized by means of a global representation bypassing the identification of the constituent objects. In [5] the colour and grey-level distribution in the image is used to separate different scene categories. In [13], [14] Gabor-Filters and in [15], [6] Fourier analysis methods are used to extract edge information from the image. A comparison of different context feature descriptions can be found in [12].

In this work, context is defined as a set of low-level image features describing the scene structure as a single, coherent entity. The impact of the context information to our probabilistic object recognition procedure is modelled using parameterised priors for a given set of object categories. The parameters of these priors are determined in a training stage. In this way, the model can account for the changes in the surrounding of the vehicle and hence induce modifications in the object detection process as e.g. accelerate a pedestrian detection process by focusing the attention to scene-specific (urban vs. rural domain) regions in the image. Trying to avoid the use of arbitrarily defined or heuristically motivated a-priori object knowledge in our recognition scheme, we argue that with the context-aware object prior the scene understanding process can be enriched by a measurable quantity. Using such a prior makes it possible to incorporate information into the recognition process which is very intuitive to humans. Obviously, this is almost impossible to express within a method that exclusively considers object-inherent information. The model presented in this work is especially interesting for an autonomously navigating system as such a system is operating in a constantly changing environment. By utilising the scene information, an object recognition method can adapt to some scene-specific object characteristics and therefore improve the recognition performance.

The remainder of the paper is organized as follows. Section II presents some of the theoretical background underlying the idea of context-aware object priors. The descriptor for expressing scene context and the context model will be presented in Section III. Section IV presents the experimental results before conclusions are drawn in Section V.

II. Scene Context in Images

Inspired by recent research results [13], in this work we extend our existing scene segmentation method [1] by the global feature *scene context*. To be able to pour context information into our object detection framework, first the question of how to characterise and describe a scene has to be answered. It is important to note that such a description must be independent of any object-intrinsic measure. Experiments in human scene perception [6] have shown, that the most requested criterion to classify an image is the *nativeness* of the displayed scene, i.e. to what extend the scene is structured in strong geometrical patterns originating from man-made structures. However, according to this study, the dominant criterion of humans to discriminate deliberate images is the *openness* of a scene. While spacious areas and the presence of the horizon indicate an open scene, the existence of close and high-rising structures are a sign for a low degree of openness. We argue that the openness of a scene is an adequate criterion to characterise a traffic scene. To be able to incorporate the context information into our existing framework, a probabilistic model is elaborated in the sequel.

A. Probabilistic Model

We wish to segment an image $\mathrm{G}(\mathrm{P})$, consisting of a set of pixels $\mathrm{P} = \{\mathrm{p}_1, .., \mathrm{p}_N\}$,$\mathrm{p}_i \in \mathbb{R}^2$ into figure, i.e. image points p_i belonging to an object category O and background. Taking a Bayesian perspective, a set of binary labels $\mathbf{l} = (l(\mathrm{p}_1), .., l(\mathrm{p}_N)) = (l_1, .., l_N)$ is defined, with one label $l_i \in \{O, background\}$ for each image point and posterior probability

$$P(\mathbf{l} = O|\mathbf{Y};\mathbf{M}) = \frac{P_{\mathbf{M}}(\mathbf{Y}|\mathbf{l}) \cdot P_{\mathbf{M}}(\mathbf{l})}{P_{\mathbf{M}}(\mathbf{Y})} \ . \tag{1}$$

$P(\mathbf{l}|\mathbf{Y};\mathbf{M})$ states the probability for the presence of an object at image position p_i (i.e. $l_i = O$) as a product of the likelihood term $P_{\mathbf{M}}(\mathbf{Y}|\mathbf{l})$ which evaluates the labeling $\mathbf{l}$ with respect to observations $\mathbf{Y}$ and the object prior $P_{\mathbf{M}}(\mathbf{l})$. $\mathbf{M}$ states the object model parameter vector of a specific object category and must be known to the system beforehand.

An important property in Equation (1) is the fact, that observations $\mathbf{Y}$ are referred to the entire image. This implicates that in statistical approaches for object recognition the high complexity of $\mathbf{Y}$ makes the modeling of $\mathbf{M}$ extremely difficult.

Most current object recognition schemes manage this complexity problem by assuming that the regions surrounding the object can be modelled independent with respect to the object presence, i.e. there exists no dependency between the scene background and the object contained in the scene. This assumption changes Equation (1) to

$$P(\mathbf{l} = O|\mathbf{Y};\mathbf{M}) \approx \frac{P_{\mathbf{M}_L}(\mathbf{Y}|\mathbf{l}) \cdot P(\mathbf{l})}{P_{\mathbf{M}_L}(\mathbf{Y})} \ . \tag{2}$$

Now, the *object-centered* posterior probability for the presence of an object is exclusively parametrised by $\mathbf{M}_L$. $\mathbf{M}_L$ states a set of local object parameters, ideally describing the local appearance and specific properties of an object in the image. Equation (2) formalises the main principle underlying the classic approach for object recognition: *the only image features that are relevant for the detection of an object at one spatial location are the features that potentially belong to the object and not to the background* [13].

Following this paradigm, in our approach object knowledge has been represented solely by a set of descriptors expressing the local appearance of salient object parts. To account for the geometric relationships among object parts, a sparse structural prior over part configurations of a specific object category has been designed. The background was thus treated as a collection of distractors but not as an entity that also conveys information about the object identity. We argue that the incorporation of object-specific scene context into our scene segmentation framework can drastically improve the process as (i) insufficient intrinsic object information can be augmented with and (ii) the exhaustive exploration of a large search space corresponding to different object models, locations and scales can be reduced by using scene context as an indicator of object presence and properties.

Scene context can be incorporated into our existing Bayesian formulation by splitting the parameter vector $\mathbf{M}$ from Equation (1) into a part $\mathbf{M}_L$ that evaluates the local object properties (see Equation (2)) and a part $\mathbf{M}_C$ that describes the complementary parts of the image, namely the background of the object. The posterior object probability can then be written

$$P(\mathbf{l} = O|\mathbf{Y};\mathbf{M}_L;\mathbf{M}_C) = \frac{P_{\{\mathbf{M}_L|\mathbf{M}_C\}}(\mathbf{Y}|\mathbf{l}) \cdot P_{\mathbf{M}_C}(\mathbf{l})}{P_{\{\mathbf{M}_L;\mathbf{M}_C\}}(\mathbf{Y})} \ . \tag{3}$$

By considering the scene context in the likelihood term $P_{\{\mathbf{M}_L|\mathbf{M}_C\}}(\mathbf{Y}|\mathbf{l})$ it is now possible to account for different appearances of the object as a function of the context. Although of great interest, in this paper we will focus on the study of the second factor $P_{\mathbf{M}_C}(\mathbf{l})$ which has received much less attention in computational vision and object recognition. While, strictly speaking, this approach violates the fundamental Bayesian principle that priors should not be estimated from data, in practice it leads to more sensible solutions than setting priors arbitrarily or using priors that are mainly motivated by heuristic assumptions. Choosing the parameter set $\mathbf{M}_C$ which maximizes $P_{\mathbf{M}_C}(\mathbf{l})$ will originate a prior which is inferred from the image data favouring the segmentation of an object in an environment that is typical for the specific object category.

B. Object Category

Next to the actual object category c_O, in this work we are further interested in the category c_S of the scene the object is contained in. For a given object category this information can then be exploited to make predictions about the scene-specific behaviour of an object. In the road-traffic domain this could e.g. be used for the behaviour prediction of pedestrians: in urban environments pedestrians are more likely to appear across the entire road while in rural areas they are more likely to remain on the footpath on the side of the road. For the purpose of this work an object is defined by

$$O = \{c_O, c_S\} \tag{4}$$

with

- c_O stating the object category. We annotate an object to one of the three categories:
  - 'car',
  - 'bicycle',
  - 'pedestrian' and
- c_S stating the scene category. There exist three different categories annotating the *openness* of the environment:
  - 'open' describes the scene as an open, poorly structured area, e.g. motorways or roads with multiple lanes,
  - 'semi-open' describes more structured and developed traffic areas as e.g. rural roads,
  - 'closed' describes highly developed and structured areas as e.g. urban and inner-city areas.

III. Context Descriptor and Scene Analogy

As already mentioned in Section II, structural information plays an important role in human scene perception. Motivated by this finding, we are now left with the task to find an efficient and meaningful set of image features to represent the structure of a scene. Depending on the specific configuration of this context feature it should be possible to separate an image into different scene categories.
In this work, the features describing the scene context are expected to quantify the magnitude and orientation of edges in different image resolutions to get an exhaustive description of the scene with fine and coarse scene structures. We used the filter responses of a *steerable pyramid* [8] which is a multi-scale, orientation sensitive gradient filter.

A. Context Description using Steerable Pyramids

Formally, the decomposition of an image $\mathrm{G}(\mathrm{P})$ by a steerable pyramid with n_{sc} scales and n_{or} orientations corresponds to a transformation $\mathscr{F}$ with an output of $K = (n_{sc} \cdot n_{or})$ filter responses $f_k(\mathrm{G}(\mathrm{P}))$

$$\mathscr{F}\{\mathrm{G}(\mathrm{P}), n_{sc}, n_{or}\} = \{f_k(\mathrm{G}(\mathrm{P}))\}_{k=1,\ldots,K} \,. \tag{5}$$

In this work the description of the scene context is based on the filter output of a steerable pyramid with $n_{or} = 6$ orientations and $n_{sc} = 4$ different scales. To describe the context features obtained by transformation $\mathscr{F}$, in the sequel we use a simplified notation[1] defined by

$$\mathbf{M}(\mathrm{G}(\mathrm{P}), k) = \mathscr{L}\{|\mathscr{F}\{\mathrm{G}(\mathrm{P}), n_{sc}, n_{or}\}|\} \,, \tag{6}$$

which assigns the magnitudes of the $K = 24$ filter responses to each image point.
Concerning $\mathbf{M}(\mathrm{G}(\mathrm{P}), K)$, there exists a strong relationship between the category of a scene and the filter responses of the steerable pyramid. Figure 2 shows the averaged magnitudes of the filter responses for the two scene categories *open* and *closed* based on a training data set of 100 images each. In

Fig. 1. A collection of images used to train the respective scene category. **Top:** images assigned to scene category *closed* **Middle:** images assigned to scene category *semi-open* **Bottom:** images assigned to scene category *open*.

Figure 1 some of the training images are shown. It can be clearly seen that numerous edges of different orientations dominate the upper half of images annotated *closed* (e.g. caused by buildings, street signs or traffic lights). In contrast, images assigned to the scene category *open* often emphasize a horizon line, sign-postings or kerbstones as strong and characterizing structures.

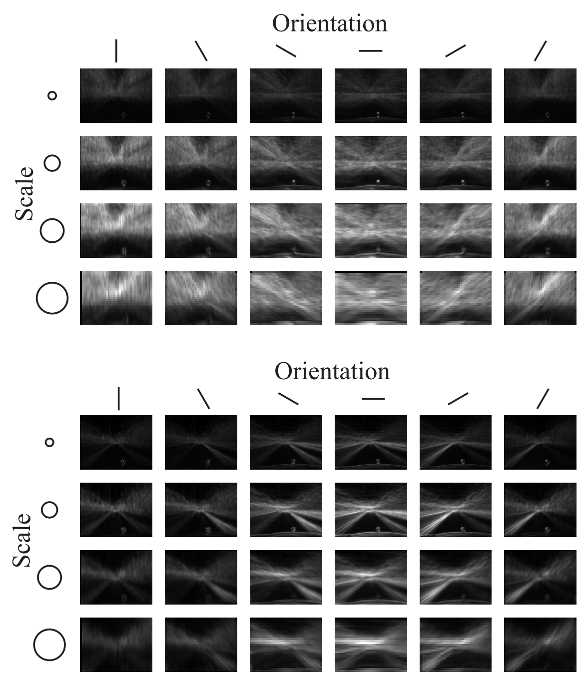

Fig. 2. Feature vector $\mathbf{M}(\mathrm{G}(\mathrm{P}), k)$ for scene category *closed* (top) and *open* (bottom).

B. Feature Vector Compression

The resulting context descriptor is very complex. To make the method computationally feasible, the descriptor is compressed and redundant information is discarded. It can

[1] symbolically, the change in notation is expressed by $\mathscr{L}(\cdot)$

be seen in Figure 2 that not only orientation and magnitude of the filter response but also the location bears characteristic scene information. To preserve the expressiveness of the descriptor it is of great importance to maintain this quantity in the compression process.
In a first step the complexity of $\mathbf{M}(\mathrm{G}(\mathrm{P}),k)$ is reduced by a sub-sampling process according to

$$\mathbf{M}^{sub}(\mathbf{s},k) = \sum_{\mathrm{P}\in\mathbf{B}(\mathbf{s},k)} \frac{\mathbf{M}(\mathrm{G}(\mathrm{P}),k)}{n_B}\,. \tag{7}$$

For a fixed k, all matrix entries of $\mathbf{M}(\mathrm{G}(\mathrm{P}),k)$ within the isotropic block size $\mathbf{B}$ are merged to one super-pixel with coordinates $\mathbf{s} = (s_x, s_y)^T$ as illustrated in Figure 3. Considering the original image proportions, every sub-sampled filter response $\mathbf{M}^{sub}(\mathbf{s},k)$ is composed of $8 \cdot 6 = 48$ super-pixels. Rearranging all of these super-pixels to a more compact view leads to a 1152-dimensional vector. This vector is further compressed using principle component analysis (PCA). In this work the $n_{pc} = 100$ components of the feature vector with largest Eigenvalue[2] have been identified as sufficient to describe the scene context and will be subsequently noted as context descriptor $\mathbf{M}_C$.

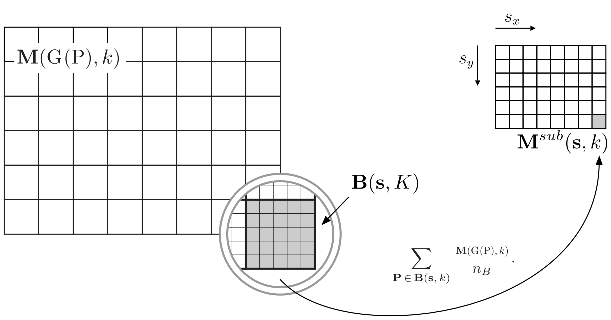

Fig. 3. Sub-sampling process of descriptor $\mathbf{M}(\mathrm{G}(\mathrm{P}),k)$. The obtained context descriptor $\mathbf{M}^{sub}(\mathbf{s},k)$ is composed of 48 super-pixels resulting in a 1152-dimensional vector.

C. Modelling Scene Context

With the single components of the context descriptor being stochastically independent and assuming that they are normally[3] distributed, the context model for a scene category c_S is formally expressed by

$$p(\mathbf{M}_C|c_S) \approx \prod_{j=1}^{h=n_{pc}} p(\mathbf{M}_{C,j}|c_S)\,, \tag{8}$$

with

$$p(\mathbf{M}_{C,j}|c_S) = \mathcal{N}(\mu_j^{C_S}, \sigma_j^{C_S})\,. \tag{9}$$

The model parameters $\mathbf{M}_{C,j} = (\mu_j^{C_S}, \sigma_j^{C_S})$ have been determined in a training stage which will be presented in more detail in the next section.

[2]the dimensionality reduction of $\mathbf{M}^{sub}(\mathbf{s},k)$ retained $97,16\%$ of the total variance

[3]The hypothesis was refined making a χ^2-test of goodness of fit at a significance level of $\alpha = 0,05$.

IV. Training and Evaluation

To be able to compute context-aware object priors, a training stage is required in which the system learns the relationship between the contextual features presented in Section III and the scene-specific object properties. As mentioned in Section II-A, we are interested in $P_{\mathbf{M}_C}(\mathbf{l} = O)$ with $O = \{c_O, c_S\}$, i.e. the probability of some object property that can be inferred based on the contextual parameter set $\mathbf{M}_C$. Objects can be annotated $c_O = \{$car, bicycle, pedestrian$\}$ stating the object category and $c_S = \{$open, semi-open, closed$\}$ stating the scene category respectively.
According to this notation, the prior probability from Equation (3) can be separated into

$$P_{\mathbf{M}_C}(\mathbf{l}) = P_{\mathbf{M}_C}(\mathrm{P}|c_O,c_S) \cdot P_{\mathbf{M}_C}(c_O \mid c_S) \cdot P_{\mathbf{M}_C}(c_S)\,, \tag{10}$$

with

- $P_{\mathbf{M}_C}(c_S)$ stating the probability of a specific scene category and
- $P_{\mathbf{M}_C}(c_O \mid c_S)$ stating the probability of a specific object category and
- $P_{\mathbf{M}_C}(\mathrm{P}|c_O,c_S)$ stating the probability of image points P to belong to object category c_O and scene category c_S based on the context model expressed by $\mathbf{M}_C$.

If it is assumed that the context descriptor $\mathbf{M}_C$ conveys enough information about the identity of a scene, strong priors on the existence of some object category within this scene can be drawn.

A. Scene Classification

Above, $P_{\mathbf{M}_C}(c_S)$ is the probability of an image to belong to scene category c_S, based on the scene-specific context model parametrised with $\mathbf{M}_C$. Considering observations $\mathbf{Y}$ in the classification process results in

$$\begin{aligned} P(c_S|\mathbf{Y};\mathbf{M}_C) &= \frac{p_{\mathbf{M}_C}(\mathbf{Y}|c_S)\cdot P(c_S)}{p_{\mathbf{M}_C}(\mathbf{Y})}\,, \text{ with} \\ p_{\mathbf{M}_C}(\mathbf{Y}) &= \sum_i p_{\mathbf{M}_C}(\mathbf{Y}|c_{S_i})\cdot P(c_{S_i})\,. \end{aligned} \tag{11}$$

$P(c_S)$ states the a-priori probability for the presence of a certain scene category. In this work we presumed all scene categories to be equally likely, i.e. $P(c_S) = \frac{1}{3}$. The second term $p_{\mathbf{M}_C}(\mathbf{Y}|c_S)$ describes the likelihood, or image evidence, of scene category c_S characterised by the trained parameter configuration of $\mathbf{M}_C$. The observation set $\mathbf{Y}$ consists of the output of the steerable pyramid filter introduced in Section III-A.

B. Scene-Specific Object Classification

In Equation (10) the second term describes the probability for the occurrence of an object of category c_O conditioned on scene category c_S and $\mathbf{M}_C$. We make the simplifying assumption that the object category is only conditioned on the scene category, i.e.

$$P_{\mathbf{M}_C}(c_O \mid c_S) \approx P(c_O \mid c_S)\,. \tag{12}$$

Based on a training data set of 300 images the probability $P(c_O \mid c_S)$ was determined heuristically according to Figure 4.

	$c_O =$ car	$c_O =$ bycicle	$c_O =$ pedestrian
$c_S =$ open	80 %	15 %	5 %
$c_S =$ semi-open	50 %	25 %	25 %
$c_S =$ closed	33 %	33 %	33 %

Fig. 4. $P(c_O \mid c_S)$ for $c_O = \{$car, bicycle, pedestrian$\}$ conditioned on scene category $c_S = \{$open, semi-open, closed$\}$.

C. Scene-Specific Object Occurrence

The first term in Equation (10) states the probability of image points P to be assigned to object category O given object category c_O and scene category c_S. In a training step the characteristic spatial distribution

$$P_{\mathbf{M}_C}(\mathrm{P}|c_O, c_S) \approx P(\mathrm{P}|c_O, c_S). \tag{13}$$

of the single object categories conditioned on the scene categories has been determined. Figure 5 shows the occurrence distribution of the different object categories as a bar chart and the probability distributions modelled as a Mixture-of-Gaussians (MoG).

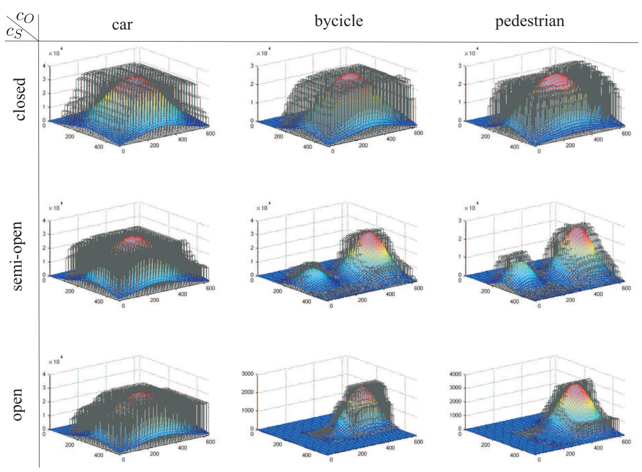

Fig. 5. Characteristic occurrence probability $P(\mathrm{P}|c_O, c_S)$ in the image plane for object categories $c_O = \{$car, bicycle, pedestrian$\}$ conditioned on scene category $c_S = \{$open, semi-open, closed$\}$. Next to the bar charts (black) also the approximated probability distributions are shown.

D. Context-Aware Object Prior

With Equation (11), Equation (12) and Equation (13) the scene-specific object prior in Equation (10) is fully defined. Now, for every position p_i in the image, the probability of object category c_O based on scene category c_S can be determined.

The previously presented context-aware object prior has been evaluated based on the results of 149 test images. In Figure 6 the classification results for the individual scene categories with respect to the dimension of the context descriptor are shown. A scene category was assumed to be classified as correct, if the probability measure was $P(c_S|\mathbf{Y};\mathbf{M}_C) > 0,95$. Furthermore, the number of misclassifications ($P(c_S|\mathbf{Y};\mathbf{M}_C) < 0,50$) is shown. For all annotations to

	True positives $P(c_S\|\mathbf{Y};\mathbf{M}_C) > 0.95$				False positives $P(c_S\|\mathbf{Y};\mathbf{M}_C) < 0.5$			
	# of principle components				# of principle components			
	100	75	50	25	100	75	50	25
$c_S =$ open	92,9 %	85,7 %	75,0 %	60,7 %	3,6 %	3,6 %	10,7 %	17,6 %
$c_S =$ semi-open	80,3 %	72,7 %	63,6 %	43,9 %	15,2 %	18,2 %	25,8 %	42,4 %
$c_S =$ closed	85,5 %	76,4 %	69,1 %	60,0 %	9,1 %	10,9 %	14,5 %	21,8 %

Fig. 6. Scene classification for a variable number of principle components.

object category *car* an average probability of $P_{\mathbf{M}_C}(\mathbf{l} = O) > 0,60$ could be guaranteed while this lower bound could be fixed for over 90% of the objects assigned to category *bicycle* or *pedestrian*. Figure 7 shows the expected occurrence maps for the individual object categories in different environments. The context-aware prior probability of an object category

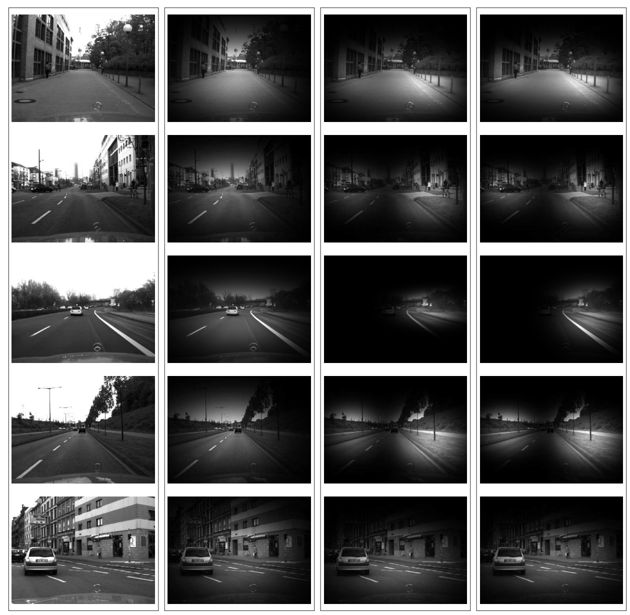

Fig. 7. In this example, the occurrence probability for a certain object category is simply a measure of local intensity contrast, i.e. regions with a low probability are black, regions with a high probability keep their original intensity value. Above, the **1st column** shows the original image, the **2nd column** shows the occurrence map of object category *car*, the **3rd column** shows the occurrence map of object category *bicycle* and the **4th column** shows the occurrence map of object category *pedestrian*.

c_O is expressed as a linear combination over all scene categories c_S. In Figure 8 the probability $P_{\mathbf{M}_C}(\mathbf{l}_{\mathbf{S}_i} = O) = \frac{1}{C}\sum_{j=1}^{C} P_{\mathbf{M}_C}(l_j = O)$ of image segment $\mathbf{S}_i = (\mathrm{p}_{i,1}, .., \mathrm{p}_{i,C})$ to belong to object category car, bicycle and pedestrian is illustrated. Object hypothesis have been generated based on the local object property *motion similarity* as described in [1].

V. Conclusion

In this contribution, a model has been presented that evaluates the strong relationships between an object and the scene it is contained in. A statistical measure, describing the

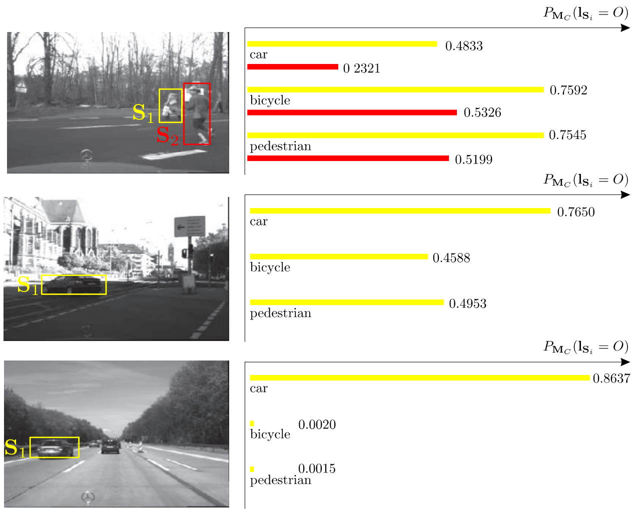

Fig. 8. **Left:** Detected objects based on the local object property *motion similarity* in three-dimensional space. **Right:** Prior probability $P_{\mathbf{M}_C}(\mathbf{l}_{\mathbf{S}_i} = O)$ of image region $\mathbf{S}_i$ to belong to object category $O = \{\text{car, bicycle, pedestrian}\}$.

influence of characteristic scene properties to object priors, has been developed. The expressiveness of the context-aware object prior has been evaluated on real image data.

In ongoing work, we expect to increase the performance of the method by further refining and extending the context descriptor. Additionally, the performance shall be increased by an exhausting training of different object and scene categories. Furthermore it is intended to fully incorporate the context-aware object prior into our existing scene segmentation framework.

ACKNOWLEDGEMENT

This work was partly supported by the Deutsche Forschungsgemeinschaft DFG within the collaborative research center 'Cognitive Automobiles'.

REFERENCES

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

[2] I. Biederman, R. J. Mezzanotte, and J. C. Rabinowitz. Scene perception: Detecting and judging objects undergoing relational violations. In *Cogintive Psychology*, number 14, pages 143–177, 1982.

[3] A. Bobick and C. Pinhanez. Using approximate models as source of contextual information for vision processin. In *ICCV 1995; Workshop on Context-Based Vision, Cambridge*, pages 13–21, 1995.

[4] D. Hoiem, A.A. Efros, and M. Hebert. Geometric context from a single image. In *IEEE International Conference on Computer Vision (ICCV'05)*, volume 1, pages 654–661, Washington, DC, USA, 2005. IEEE Computer Society.

[5] A. Oliva and P. G. Schyns. Diagnostic color blobs mediate scene recognition. In *Cogintive Psychology*, number 41, pages 176–210, 2000.

[6] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. In *International Journal of Computer Vision*, volume 3, 2001.

[7] T.H. Rockwell. Skills, judgment, and information acquisition in driving. *Human Factors in Highway Traffic Safety Research*, pages 133–164, 1972.

[8] E.P. Simoncelli and W.T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *IEEE International Conference on Image Processing*, 1995.

[9] M. Sivak. The information that drivers use: is it indeed 90*Perception*, 25(9):1081–1089, 1996.

[10] X. Song, J. Sill, Y. Abu-Mostafa, and H. Kasdan. Image recognition in context: Application to microscopic urinalysis. In *Advances in Neural Information Processing Systems*, pages 963–969, 2000.

[11] T.M. Strat and M.A. Fischler. Context-based vision: Recognizing objects using information from both 2-d and 3-d imagery. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 13, pages 1050–1065, 1991.

[12] M. Szummer and R.W. Picard. Indoor-outdoor image classification. In *IEEE International Workshop on Content-Based Access of Image and Video Databases*, 1998.

[13] A. Torralba. Contextual priming for object detection. In *International Journal of Computer Vision*, volume 52, pages 169–191, 2003.

[14] A. Torralba. Contextual influences on saliency. In *MIT Artificial Intelligence Laboratory*, volume 9, 2004.

[15] A. Torralba and P. Sinha. Statistical context priming for object detection. In *International Conference on Computer Vision*, pages 763–770, 2001.