Object-level Priors for Stixel Generation

Marius Cordts[1,2], Lukas Schneider[1], Markus Enzweiler[1],
Uwe Franke[1], and Stefan Roth[2]

[1] Environment Perception, Daimler R&D, Sindelfingen, Germany
[2] Department of Computer Science, TU Darmstadt, Germany
marius.cordts@daimler.com

Abstract. This paper presents a stereo vision-based scene model for traffic scenarios. Our approach effectively couples bottom-up image segmentation with object-level knowledge in a sound probabilistic fashion. The relevant scene structure, i.e. obstacles and freespace, is encoded using individual Stixels as building blocks that are computed bottom-up from dense disparity images. We present a principled way to additionally integrate top-down prior information about object location and shape that arises from independent system modules, ranging from geometric cues up to highly confident object detections. This results in an efficient exploration of orthogonal image-based cues, such as disparity and gray-level intensity data, combined in a consistent scene representation. The overall segmentation problem is modeled as a Markov Random Field and solved efficiently through Dynamic Programming.
We demonstrate superior segmentation accuracy compared to state-of-the-art superpixel algorithms regarding obstacles and freespace in the scene, evaluated on a large dataset captured in real-world traffic.

1 Introduction

Visual scene understanding is a key problem for autonomous driving and robotics. Especially the knowledge of obstacles that limit the available freespace is cru-

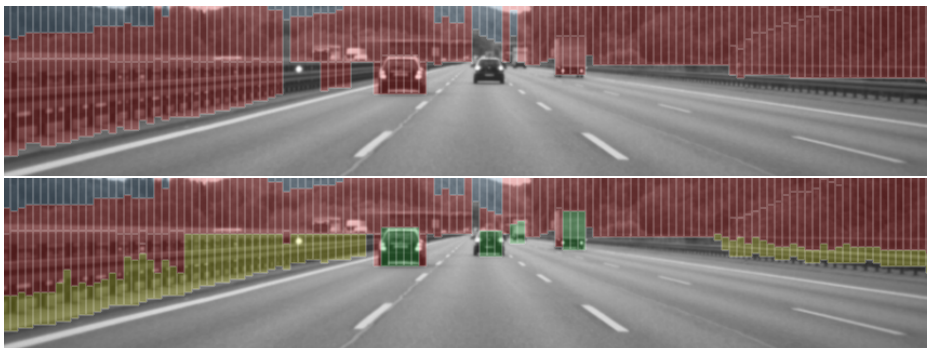

Fig. 1: Original Stixel world (top) and our extension (bottom). Stixel classes are freespace (transp.), obstacle (red), sky (blue), vehicle (green), guard rail (yellow).

cial for navigation and collision avoidance. This segmentation task was tackled using dense stereo imaging by Badino et al. [4], and resulted in the so-called Stixel world. Subsequently, the model was extended by Pfeiffer et al. [22] to a full image segmentation providing a compact medium-level scene representation accurately capturing multiple depth-layers of objects. These superpixels reduce the complexity for subsequent image processing tasks and are successfully used for numerous applications, such as semantic segmentation [23], object detection [5, 12], mapping [21] or segmentation of dynamic objects [13].

However, Stixels are solely based on dense stereo and a strongly simplifying world model with a nearly planar road surface and perpendicular obstacles. Thus, whenever depth measurements are noisy or the world model is violated, Stixels are prone to errors. As can be seen in Fig. 1 top, the car in the center lane, the distant truck and the guard rails are not accurately segmented. In contrast to the bottom-up Stixel segmentation, top-down object detectors do not suffer from the mentioned limitations but are specific for one object class, e.g. vehicle, pedestrian or guard rail, and do not provide a generic scene representation.

The main contribution of this work is to show a principled way to incorporate such top-down prior knowledge into the Stixel generation combining the strengths of both methods. We follow a probabilistic approach that allows to find the optimal solution of an extended world model. The additional information not only improves the representation of the detected object classes, but also influences the inference of other parts in the scene, e.g. the freespace. We evaluate our approach in a highway scenario using detectors for vehicles and guard rails, see Fig. 1 for an exemplary output. From a practical point of view, the resulting Stixel world unifies various sources of information and provides a clean, simple and consistent interface for subsequent processing stages.

1.1 Related Work

We see four major categories of publications related to our work. The first contains algorithms for unsupervised image segmentation [1, 3, 15]. Such methods do not use any semantics and aim for a generic representation of the scene with reduced complexity. Most algorithms are based on appearance only, but there are some that also utilize stereo [26, 30]. The Stixel world [22] as introduced above and extended in this publication is naturally closest related to our work.

The second category comprises top-down methods such as detectors for objects [8,10,18,27] or geometric shapes [9,17]. These detectors often show excellent performance, but can only be applied to specific object classes and do not contribute to an understanding of the remaining scene. In this work, we fill this gap and show how to leverage detectors for improving the generic scene model.

The third category covers the task of semantic segmentation, i.e. each segment is associated with a class label. Such methods either operate on a pixel level [19, 25] or use superpixels as smallest considered unit [6, 7, 16]. Although our proposed algorithm provides an image segmentation with associated class labels, we do not claim to perform semantic segmentation. Our labels are restricted to be ground, sky, generic obstacle or those that object detectors provide, which

is not sufficient for a full semantic labeling. However, the proposed Stixel world is expected to serve well as superpixels for a subsequent labeling step [23].

The fourth and last category uses methods of the second category for semantic segmentation [2,20,29]. From the methodology point of view, we see our work closest related to publications in this group, using a probabilistic approach for integrating top-down information in a bottom-up task. However, those methods either reason on pixel-level [20, 29] or use superpixels as the finest element [2]. The first is generally computationally expensive and the latter cannot recover from errors already present in the superpixels. Thus, in our work we use top-down knowledge one step earlier, i.e. during the superpixel generation.

2 The Stixel World

In this Section, we describe the Stixel computation as introduced in [22]. However, we reformulate the Stixel world using a Markov Random Field (MRF) in order to integrate object-level prior knowledge in Section 3.

The Stixel world $\boldsymbol{\mathcal{S}}$ is a segmentation of an image $\mathcal{I}$ with size $w \times h$ into so-called Stixels, where each pixel is assigned to exactly one Stixel. Such a Stixel $\boldsymbol{s}_{ui} \in \boldsymbol{\mathcal{S}}$ can be seen as a superpixel, however restricted to be a vertical line segment in a certain column $u \in \{1 \dots w\}$ with bottom and top row $v_{ui}^{\mathrm{b}} \leq v_{ui}^{\mathrm{t}}$. If the image is horizontally sub-sampled, Stixels become the rectangles visualized in Fig. 1. The enumeration i of Stixels within a column is such that $v_{ui}^{\mathrm{t}}+1 = v_{u,i+1}^{\mathrm{b}}$. According to the assumed world model, a scene is composed of a ground g, perpendicular objects o and the sky s. Thus, each Stixel is associated with a class $c_{ui} \in \{\mathrm{g}, \mathrm{o}, \mathrm{s}\}$ and a disparity $d_{ui} \in \mathbb{R}_{\geq 0}$. The latter is discretized, not defined for the ground, represents the disparity of an object and is zero for the sky. Together, a Stixel $\boldsymbol{s}_{ui}$ is sufficiently described by the tuple $\boldsymbol{s}_{ui} = \left(v_{ui}^{\mathrm{b}}, v_{ui}^{\mathrm{t}}, c_{ui}, d_{ui}\right)$.

2.1 Probabilistic reformulation

Treating all columns independently, searched Stixels $\boldsymbol{S}_{u:}$ in column u are interpreted as random variables $\boldsymbol{S}_{u:} = (\boldsymbol{S}_{u1}, \boldsymbol{S}_{u2}, \dots, \boldsymbol{S}_{un_u})$ for a fixed number of Stixels n_u. The measured disparity image in column u is denoted as $\boldsymbol{d}_{u:}$, being the observations of random variables $\boldsymbol{D}_{u:}$. The posterior $P(\boldsymbol{S}_{u:} \mid \boldsymbol{D}_{u:})$ is defined using an MRF, depicted as a factor graph in Fig. 2. The graph provides a factorization grouped into a likelihood $\Phi(\boldsymbol{d}_{u:}, \boldsymbol{s}_{u:})$ and a prior $\Psi(\boldsymbol{s}_{u:})$, giving

$$P(\boldsymbol{S}_{u:} = \boldsymbol{s}_{u:} \mid \boldsymbol{D}_{u:} = \boldsymbol{d}_{u:}) = \frac{1}{Z} \Phi(\boldsymbol{d}_{u:}, \boldsymbol{s}_{u:})\, \Psi(\boldsymbol{s}_{u:}) \ , \tag{1}$$

where Z is the normalizing partition function. The final segmentation $\boldsymbol{s}_{ui}^{\star}$ is obtained after selecting a model $n_u^{\star}$ and solving the maximum-a-posteriori (MAP) problem

$$\boldsymbol{s}_{u:}^{\star} = \underset{\boldsymbol{s}_{u:}}{\operatorname{argmax}}\ P(\boldsymbol{S}_{u:} = \boldsymbol{s}_{u:} \mid \boldsymbol{D}_{u:} = \boldsymbol{d}_{u:}) \ . \tag{2}$$

Note that this problem can be solved efficiently using Dynamic Programming (DP), see [22]. Model selection, measurement likelihood and prior are discussed individually in the following, while omitting the column index u for readability.

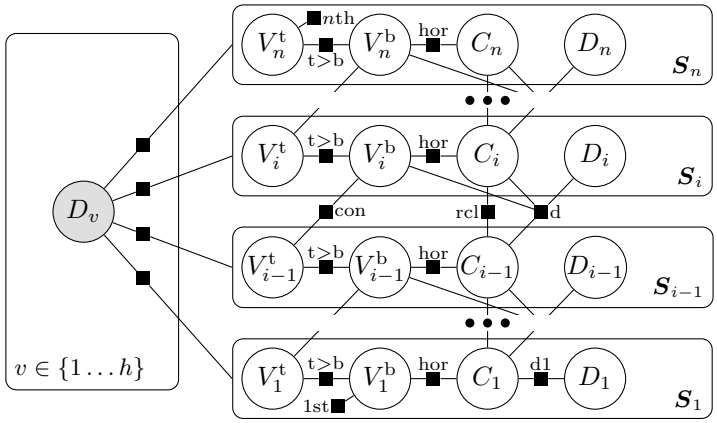

Fig. 2: Stixel world as an MRF. Each Stixel $\boldsymbol{S}_i$ (horizontal boxes) stands for the four random variables $V_i^{\mathrm{t}}, V_i^{\mathrm{b}}, C_i, D_i$ (circles). The prior distribution factorizes according to the labeled factors (black squares, descriptions given in the text). The left part contains exemplary one observed disparity D_v (shaded circle) and the connected factors of the measurement likelihood.

Model selection For selecting a model $n^\star$, first the maximum value $p_n^\star$ of the posterior $P(\boldsymbol{S}_{:} \mid \boldsymbol{D}_{:})$ is determined for each model $n \in \{1 \dots h\}$. Each value $p_n^\star$ is weighted with a model complexity prior $\exp(-\alpha n_u)$ and the model giving the maximum result is selected. Approximating the partition function Z as being constant for all models, model selection and MAP estimation are performed simultaneously in one DP sweep without explicitly computing the maxima $p_n^\star$.

Likelihood Assuming conditional independence of the disparities given the segmentation, see Fig. 2, the likelihood factorizes as

$$\Phi(\boldsymbol{d}_{:}, \boldsymbol{s}_{:}) = \prod_{v=1}^{h} \prod_{i=1}^{n} \Phi_v(d_v, \boldsymbol{s}_i) \ . \tag{3}$$

The term $\Phi_v(d_v, \boldsymbol{s}_i)$ represents the disparity measurement model that describes the probability of a value d_v for a given Stixel $\boldsymbol{s}_i$ and is non-informative, i.e. uniform, for Stixels not covering row v. For details on the likelihood see [22].

Prior The prior is modeled using the right part of the Markov random field in Fig. 2. Most important is the first order Markov assumption on Stixel-level, inducing conditional independence of a Stixel $\boldsymbol{S}_i$ and its non-neighbors given its neighbors. All factors belonging to the prior $\Psi(\boldsymbol{s}_{:})$ are introduced subsequently, see their labels in Fig. 2.

The factors $\Psi_{1\mathrm{st}}\big(v_1^{\mathrm{b}}\big)$, $\Psi_{n\mathrm{th}}(v_n^{\mathrm{t}})$, $\Psi_{\mathrm{t>b}}\big(v_i^{\mathrm{b}}, v_i^{\mathrm{t}}\big)$, $\Psi_{\mathrm{con}}\big(v_i^{\mathrm{b}}, v_{i-1}^{\mathrm{t}}\big)$ and $\Psi_{\mathrm{hor}}\big(v_i^{\mathrm{b}}, c_i\big)$ enforce a consistent segmentation: the 1st Stixel starts in row 1, the nth ends in

row h, the top row v_i^{t} is greater than the bottom v_i^{b}, Stixels are connected, i.e. $v_i^{\mathrm{b}} = v_{i-1}^{\mathrm{t}} + 1$, and there is no sky below or ground above the horizon.

The factor $\Psi_{\mathrm{rcl}}(c_i, c_{i-1})$ models the probabilities for relative class locations, e.g. an object Stixel below of a ground one is less likely than vice versa. The remaining factors $\Psi_{\mathrm{d1}}(d_i, c_i)$ and $\Psi_{\mathrm{d}}\big(d_i, v_i^{\mathrm{b}}, c_i, c_{i-1}\big)$ describe the probability distribution of disparities d_i. For $c_i = \mathrm{g}$ the disparity value is not defined and thus the distribution does not matter, whereas for $c_i = \mathrm{s}$ the probability is only non-zero for $d_i = 0$. In case of $c_i = \mathrm{o}$ the factors split into two functions as

$$\Psi_{\mathrm{d1}}(d_i, c_i = \mathrm{o}) = f_{\mathrm{blg}}(d_i, 1) \tag{4}$$

$$\Psi_{\mathrm{d}}\big(d_i, v_i^{\mathrm{b}}, c_i = \mathrm{o}, c_{i-1}\big) = f_{\mathrm{blg}}\big(d_i, v_i^{\mathrm{b}}\big)\, f_{\mathrm{grav}}\big(d_i, v_i^{\mathrm{b}}, c_{i-1}\big) \ . \tag{5}$$

Both functions use the known camera geometry to derive the expected disparity $d_{\mathrm{g}}\big(v_i^{\mathrm{b}}\big)$ of the ground in row v_i^{b}. Then a value $d_i < d_{\mathrm{g}}\big(v_i^{\mathrm{b}}\big)$ indicates an object below the ground surface, which is unlikely and captured by f_{blg}. Further, a value $d_i > d_{\mathrm{g}}\big(v_i^{\mathrm{b}}\big)$ and a preceding class $c_{i-1} = \mathrm{g}$ means that the object is flying above the ground, i.e. no gravity, which is rated with low probability by f_{grav}.

For details on the individual factors, see their corresponding probability distributions in [22].

3 Incorporating Priors

The original Stixel generation is solely based on the disparity image and computed independently for each column. However, other sources of information are often available that take into account the gray value image or multiple columns in the disparity image. This information usually applies only to a specific class and describes its rough location in the image, e.g. using a bounding box, a more complex contour or just a line indicating one end of the object.

3.1 Generic prior model

We assume that for each kind of information $j \in \{1 \dots m\}$, we have a model describing for all pixels the unnormalized probability of being a bottom or top point of the object. The model takes into account the reliability of the source of information, its importance for the Stixel generation, and also a possible uncertainty in the precise location. Such a mapping from a given contour to actual bottom and top point probabilities could either be obtained from training data or by blurring the contour. Thus, each additional input that allows for a meaningful mapping to such probability images can be used for the Stixel generation, see Fig. 3. The resulting images are denoted as $\boldsymbol{\mathcal{B}}_j, \boldsymbol{\mathcal{T}}_j \in \mathbb{R}^{w \times h}$ and are zero where there is no input prior.

3.2 Probabilistic formulation

In order to use this information for the Stixel generation, we define additional classes a_j that we treat as objects if not stated otherwise. Let $\boldsymbol{B}_{u:}, \boldsymbol{T}_{u:}$ denote the

Fig. 3: Probability images derived from the output of three different object detectors: a bounding box, a precise contour and a line. Probabilities for the true contours are encoded as intensities with blue for bottom and red for top points.

union of all bottom/top probabilities in column u, interpreted as given random variables. Then Eq. (2) is modified to

$$\boldsymbol{s}^{\star}_{u:} = \underset{\boldsymbol{s}_{u:}}{\operatorname{argmax}}\ P(\boldsymbol{S}_{u:} = \boldsymbol{s}_{u:} \mid \boldsymbol{D}_{u:} = \boldsymbol{d}_{u:}, \boldsymbol{B}_{u:} = \boldsymbol{b}_{u:}, \boldsymbol{T}_{u:} = \boldsymbol{t}_{u:}) \ . \tag{6}$$

We leave the likelihood unchanged and modify the prior using $\boldsymbol{b}_{u:}$ and $\boldsymbol{t}_{u:}$. First, the factor $\Psi_{\mathrm{rcl}}(c_i, c_{i-1})$ is extended to model the relative location probabilities of the introduced object sub-classes. All additional classes a_j are forced to be on the ground, i.e. $\Psi_{\mathrm{rcl}}(c_i = a_j, c_{i-1} = \mathrm{g}) = 1$ and $\Psi_{\mathrm{rcl}}(c_i = a_j, c_{i-1} \neq \mathrm{g}) = 0$. The case $c_{i-1} = a_j$ is treated as $c_{i-1} = \mathrm{o}$. Second, we introduce an additional factor $\Psi_{\mathrm{pi}}\big(v_i^{\mathrm{t}}, v_i^{\mathrm{b}}, c_i\big)$ being the only part where we make use of the bottom and top point probability images. For c_i being one of the standard classes $\mathrm{g}, \mathrm{o}, \mathrm{s}$, the factor Ψ_{pi} is 1. If c_i is one of the additional classes a_j, it holds

$$\Psi_{\mathrm{pi}}\big(v_i^{\mathrm{t}}, v_i^{\mathrm{b}}, c_i = a_j\big) = b_j\big(v_i^{\mathrm{b}}\big)\, t_j\big(v_i^{\mathrm{t}}\big) \ , \tag{7}$$

where $b_j(v)$ and $t_j(v)$ denote the values of bottom/top point probabilities for class a_j in row v. Where detections are present, b_j and t_j are typically greater than 1, rating the class a_j more likely than the standard classes. If b_j or t_j are 0, i.e. no matching detection, the factor Ψ_{pi} is 0 and hence also the probability for a_j. Third, we add the factor $\Psi_{\mathrm{ht}}\big(v_i^{\mathrm{t}}, v_i^{\mathrm{b}}, c_i, d_i\big)$ that captures expectations on the height of an object a_j and evaluates to 1 for valid heights and 0 otherwise. The height in world coordinates can be computed using the given arguments as well as known camera parameters. The described adaptations add only minor computational overhead and still allow for an efficient solution using DP.

4 Experimental Results

In the experimental section, we apply the extended Stixel model to a highway scenario using two external sources of information. Significant improvement is shown in two key properties: segmentation accuracy (Section 4.3) and freespace estimation (Section 4.5). Results are compared to the original Stixel formulation [22]. Additionally, the influence of our method on the precision of vehicle Stixels is evaluated (Section 4.4). Throughout all experiments, we use identical parameters for our approach and only perform the modifications explained above. Our method increases the runtime of the Stixel computation by 14 %.

Fig. 4: Exemplary detections and resulting bottom/top point probabilities.

4.1 Priors

In this paper we focus on the methodology for incorporating external priors in the Stixel computation. This general idea is independent of the particular information used and can in principle extend to arbitrary object classes. For the experimental evaluation, we utilize detectors for vehicles and guard rails, both highly relevant for autonomous driving on highways. The vehicle detector contributes to all three conducted experiments, see Sections 4.3 to 4.5, whereas the guard rails mainly influence the freespace evaluation in Section 4.5. Possible detector outputs and the resulting bottom/top point probabilities are visualized in Fig. 4. Note how this knowledge helps to improve the Stixel world, see Fig. 1.

Vehicles Out of a multitude of proposed vehicle detectors [27], we opted for a two-step system that couples high detection performance at large distances with real-time computational efficiency. In particular, we rely on a very fast vehicle detector, i.e. a Viola-Jones cascade detector [28], to create regions-of-interest for a subsequent strong set of Mixture-of-Experts classifiers using local receptive field features (LRF) [11]. In doing so, the output of the vehicle detector describes the rough location of the vehicle's rear side in the image and its associated confidence value p_{conf}. To incorporate prior knowledge about a vehicle's shape, both the bottom and the top boundary are blurred using a Gaussian truncated at 3σ. The bottom one is centered on the boundary and has a rather small variance. Since the top of a car is less accurately described by a line, we center the Gaussian along a downwards parabola and increase the variance from the box center towards its border, for an example see Fig. 3 left. The maximum value of both probability images is set to $p_{\text{conf}} \exp(\alpha_{\text{v}})$. Eventually, we model the height of vehicles between 0.5 m and 5 m using the factor Ψ_{ht}, see Section 3.2.

Guard rails As guard rail detector, we use parts of [24], based on geometry and appearance. First, the most prominent lines are found using a Hough transform on the gradient image, while restricting the search space to lines matching the expected slope. Only pixels with height values in the range of interest are considered. Second, lines whose disparity does not decrease linearly are discarded. The remaining lines are the guard rail detections and are blurred using a vertical Gaussian to model the top point probabilities with a maximum value of $\exp(\alpha_{\text{gr}})$. The height of guard rails is restricted to be between 0.4 m and 1.5 m.

Fig. 5: Example from dataset containing manual annotations (random colors).

4.2 Dataset

For evaluation, we captured a stereo sequence of 2000 frames on a German highway. Non-occluded vehicles up to very large distances are manually annotated with pixel-accuracy. In addition, the first objects limiting the driving corridor, mainly guard rails, are annotated with pixel-accuracy in every tenth frame. Eventually, occluded and approaching vehicles are annotated as "ignore" with bounding box precision. For an example see Fig. 5.

4.3 Segmentation Accuracy

One key requirement of the Stixel world is to represent the scene and contained objects accurately. To evaluate this aspect, we focus on non-occluded vehicles, since they are the most relevant objects on highways. In the conducted experiment, our method is compared to state-of-the-art superpixel methods with similar runtime, namely SLIC [1] and graph-based image segmentation (GBIS) [15].

To evaluate the algorithms from a semantic point-of-view, we assign to each generated superpixel the majority ground truth label of all covered pixels. Thus, we obtain the upper performance limits of all possible systems for semantic segmentation based on these superpixels. The average number of superpixels that are needed to represent an object serves as a measure of the segmentation's complexity. To evaluate its accuracy, we do not use the PASCAL VOC intersection-over-union (IU) [14], since this measure is dominated by foreground objects and vehicles in larger distances have only little impact. Instead, we answer the question of how well an object can be represented *on average* by computing the IU for each vehicle individually and averaging the results, see Fig. 6 left. Further, we investigate *how many* objects are accurately described via thresholding the IU at 0.5 and determining the detection rate, see Fig. 6 right. To be independent of the parameterization of the algorithms, evaluation is performed for a variation of the most relevant parameters (light markers) taken from $[1, 15, 22]$ and the upper left part of the convex hull (solid line) is used for comparison of the methods.

The results show that the proposed method significantly improves the performance of the Stixel segmentation and outperforms other state-of-the-art baselines. Especially vehicles in large distances are better segmented due to the information provided by the detector. Note that our method can directly benefit from stronger or additional detectors, whereas an incorporation into SLIC or GBIS is not straightforward.

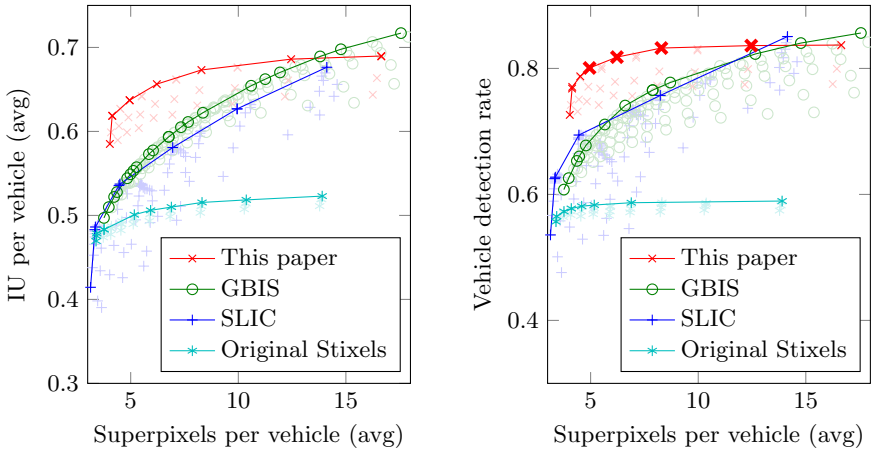

Fig. 6: Our method compared to three baselines in terms of segmentation accuracy over segmentation complexity for the class vehicle. The latter is expressed by the average number of superpixels per object. Accuracy is compared by providing the upper limits for any system based on these superpixels using the average intersection-over-union (IU) per object (left) and the detection rate (right). Each marker stands for one parameter set and solid lines connect the best performing.

4.4 Precision

The strength of our approach is the probabilistic integration of detector knowledge into the scene model. In doing so, the whole scene structure in an image column is jointly inferred and unlikely constellations such as vehicles above the ground or inappropriately sized are captured. Thus, erroneous priors due to false positive detections can be suppressed. Further, conflicts due to overlapping detections between objects of the same or different classes are optimally solved, referring to our scene model. These advantages manifest in a high precision of vehicle Stixels. For evaluation, we use the parameter sets belonging to the four red points highlighted in Fig. 6 right. For each setting, we measure the influence of the scene model's main parameter regarding vehicles, i.e. our confidence in the vehicle detector α_{v}, see Section 4.1. We evaluate the detection rate as defined above and the precision of Stixels with class vehicle. Such a Stixel is considered a true or false positive depending on its covered pixels and their ground truth labels. As baseline serves a more trivial approach for integrating object bounding boxes into the original Stixel world. Here, the obtained segmentation is post-processed to match the given boxes by adding or splitting Stixels where needed and label all Stixels within a box as vehicle.

We significantly improve the precision compared to this baseline without reducing the detection rate, see Fig. 7. Due to modeling the uncertainty of the object's location within a given bounding box, the detection rate is even increased slightly. When weakening the coupling of the detectors by decreasing

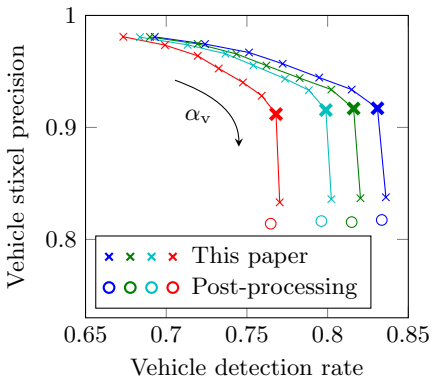

Fig. 7: Influence of detector weight α_v on detection rate (defined above) and precision of vehicle Stixels. A post-processing which forces Stixels to match detections serves as baseline. Highlighted markers correspond to those in Fig. 6.

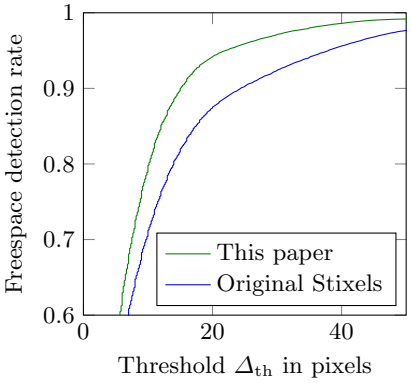

Fig. 8: Freespace detection rate of original Stixels compared to our proposal. For each Stixel column, the freespace counts as detected, if the deviation of estimation to ground truth is within the range Δ_{th}.

α_v, the precision improves even more, however at the cost of a lower detection rate.

4.5 Freespace estimation

For each Stixel column, we extract the row v delimiting the freespace from ground truth. This row is compared to the bottom of the first detected obstacle Stixel from the baseline implementation and our approach, see the transparent areas in Fig. 1. Thresholding this difference at Δ_{th} allows measuring the freespace detection rate, see Fig. 8. Our proposed method outperforms the baseline, even though we do not explicitly influence the freespace estimation. However, due to the joint inference of the whole scene, better detections of delimiting objects, i.e. vehicles and guard rails, helps to obtain an overall improved segmentation.

5 Conclusion

This work presented a principled method to integrate top-down object-level priors into bottom-up Stixel segmentation. Our approach outperformed state-of-the-art in terms of segmentation accuracy and freespace estimation at real-time speeds. For future applications, we expect our model to generalize well to additional classes of information beyond the ones presented in this paper. In addition, our approach provides powerful superpixels for semantic segmentation systems used for rural or urban traffic scenes. Ultimately, this enables to recover a much stronger understanding and interpretation of complex dynamic scenes.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012)
2. Arbeláez, P., Hariharan, B., Gu, C.: Semantic segmentation using regions and parts. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
3. Arbeláez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (2011)
4. Badino, H., Franke, U., Pfeiffer, D.: The Stixel world - a compact medium level representation of the 3D world. In: DAGM Symposium (2009)
5. Benenson, R., Mathias, M., Timofte, R., Van Gool, L.: Pedestrian detection at 100 frames per second. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
6. Carreira, J., Caseiro, R., Batista, J., Sminchisescu, C.: Semantic segmentation with second-order pooling. In: European Conference on Computer Vision (2012)
7. Dann, C., Gehler, P., Roth, S., Nowozin, S.: Pottics - the Potts topic model for semantic image segmentation. Pattern Recognition (2012)
8. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: an evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012)
9. Duda, R., Hart, P.: Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM 15(1) (1972)
10. Enzweiler, M., Gavrila, D.M.: Monocular pedestrian detection: survey and experiments. IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009)
11. Enzweiler, M., Gavrila, D.M.: A multi-level mixture-of-experts framework for pedestrian classification. IEEE Transactions on Image Processing 20(10) (2011)
12. Enzweiler, M., Hummel, M., Pfeiffer, D., Franke, U.: Efficient Stixel-based object recognition. In: IEEE Intelligent Vehicles Symposium (2012)
13. Erbs, F., Schwarz, B., Franke, U.: From Stixels to objects - a conditional random field based approach. In: IEEE Intelligent Vehicles Symposium (2013)
14. Everingham, M., Gool, L.V., Williams, C.K.I., Winn, J., Zisserman, A.: The Pascal Visual Object Classes (VOC) challenge. International Journal of Computer Vision 88 (2010)
15. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision 59 (2004)
16. Fulkerson, B., Vedaldi, A., Soatto, S.: Class segmentation and object localization with superpixel neighborhoods. In: International Conference on Computer Vision (2009)
17. Gavrila, D.M.: A Bayesian, exemplar-based approach to hierarchical shape matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007)
18. Jain, A., Duin, R., Mao, J.: Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1) (2000)
19. Ladický, L., Sturgess, P., Russell, C., Sengupta, S., Bastanlar, Y., Clocksin, W., Torr, P.H.S.: Joint optimisation for object class segmentation and dense stereo reconstruction. In: British Machine Vision Conference (2010)
20. Ladický, ., Sturgess, P., Alahari, K., Russell, C., Torr, P.H.S.: What, where and how many? Combining object detectors and CRFs. In: European Conference on Computer Vision. Springer (2010)

21. Muffert, M., Schneider, N., Franke, U.: Stix-Fusion: a probabilistic Stixel integration technique. In: Canadian Conference on Computer and Robot Vision (2014)
22. Pfeiffer, D., Franke, U.: Towards a global optimal multi-layer Stixel representation of dense 3D data. In: British Machine Vision Conference (2011)
23. Scharwächter, T., Enzweiler, M., Franke, U., Roth, S.: Efficient multi-cue scene segmentation. In: German Conference on Pattern Recognition (2013)
24. Scharwächter, T., Schuler, M., Franke, U.: Visual guard rail detection for advanced highway assistance systems. In: IEEE Intelligent Vehicles Symposium (2014)
25. Shotton, J., Winn, J., Rother, C., Criminisi, A.: TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. International Journal of Computer Vision (2009)
26. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: European Conference on Computer Vision (2012)
27. Sun, Z., Bebis, G., Miller, R.: On-road vehicle detection: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2006)
28. Viola, P., Jones, M.J.: Robust real-time object detection. International Journal of Computer Vision 4 (2001)
29. Wojek, C., Schiele, B.: A dynamic conditional random field model for joint labeling of object and scene classes. European Conference on Computer Vision (2008)
30. Zhang, J., Kan, C., Schwing, A.G., Urtasun, R.: Estimating the 3D layout of indoor scenes and its clutter from depth sensors. In: International Conference on Computer Vision (2013)