Stix-Fusion: A Probabilistic Stixel Integration Technique

Maximilian Muffert
Daimler AG
Sindelfingen, Germany
maximilian.muffert@daimler.com

Nicolai Schneider
IT-Designers GmbH
Esslingen, Germany
stz.schneider@daimler.com

Uwe Franke
Daimler AG
Sindelfingen, Germany
uwe.franke@daimler.com

***Abstract*—In summer 2013, a Mercedes S-Class drove completely autonomously for about 100 km from Mannheim to Pforzheim, Germany, using only close-to-production sensors. In this project, called *Mercedes Benz Intelligent Drive*, stereo vision was one of the main sensing components. For the representation of free space and obstacles we relied on the so called Stixel World, a generic 3D intermediate representation which is computed from dense disparity images.**

In spite of the high performance of the Stixel World in most common traffic scenes, the availability of this technique is limited. For instance under adverse weather, rain or even spray water on the windshield results in erroneous disparity images which generate false Stixel results. This can lead to undesired behavior of autonomous vehicles.

Our goal is to use the Stixel World for a robust free space estimation and a reliable obstacle detection even during difficult weather conditions.

In this paper, we meet this challenge and *fuse* the Stixels incrementally into a reference grid map. Our new approach is formulated in a Bayesian manner and is based on existence estimation methods. We evaluate our new technique on a manually labeled database with emphasis on bad weather scenarios. The number of structures which are detected mistakenly within free space areas is reduced by a factor of two whereas the detection rate of obstacles increases at the same time.

***Keywords*-stereo vision; temporal Stixel World integration; occupancy grid mapping; existence estimation**

I. Introduction

Autonomous driving is one of the most active fields of research in computer vision and robotics. Projects such as the Google self driving car [1], AnnieWAY [2] or the Mercedes Benz Intelligent Drive [3] should be mentioned in this context. For driving autonomously, the immediate environment of the ego vehicle has to be detected at all times to allow on-line navigation and path planning. Beyond that, modern driver assistance systems like lane change systems or collision avoidance systems [4] also benefit from detailed and high accurate environment models. For both applications, it is essential to minimize the number of obstacles which are detected mistakenly within free space areas. In this context, these obstacles are called false positives (fp) or *phantoms*.

A proven method for the reconstruction of the 3D environment is stereo vision [3], [5]. Today, (S)emi-(G)lobal-(M)atching [6] is a popular algorithm to estimate disparity images.

It allows to compute a depth measurement for almost every pixel in real-time [7] (c.f. Fig. 1). However, using dense stereo also results in large amounts of 3D data that have to be processed.

To reduce the computational burden for further steps, we rely on the so called multi-layered Stixel World [8] which is computed from dense disparity images (c.f. Fig. 1). In the sense of super pixels, the Stixel World efficiently describes static obstacles as well as free space information.

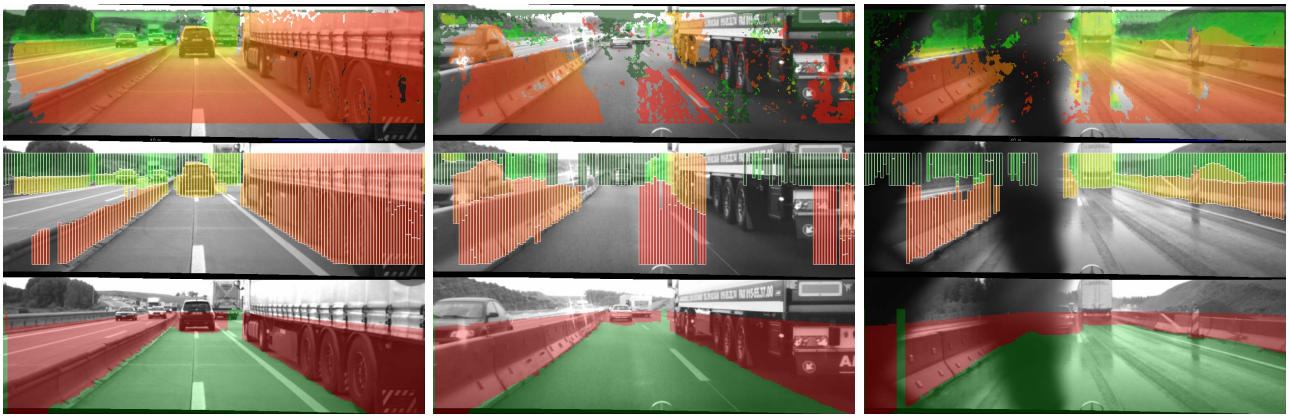
The quality of the Stixel World largely depends on the quality of the disparity image. Mistakes made during stereo matching lead to errors in the Stixel reconstruction. For example, Fig. 1(b) and Fig. 1(c) show two challenging highway scenes with heavy rain. One can clearly see *phantoms* on the ground surface caused by false stereo matches. Furthermore, stereo matching does not work in occluded image parts like when the sight is blocked by wipers crossing the windshield (c.f. Fig. 1(c)).

In this work, we focus on a reliable and robust estimation of free space and occupied areas in bad weather scenes on highways. We present a new Stixel World integration technique which is based on probabilistic existence estimation methods. Over consecutive time steps, Stixels are fused into a Cartesian occupancy grid map. An inverse projection of this fusion map into the image plane allows a comprehensible representation of free space and obstacles. Note that the new approach can also be used to build global Cartesian 2D occupancy grid maps as long as the ego motion is given.

Examples of our new technique are shown in Fig. 1 and point out the main advantages: as a result of the temporal integration, we achieve a reliable and robust representation of free space (represented by green image parts) and obstacles (represented by red image parts) even in difficult situations. *Phantoms* do not occur on the ground surface (c.f. Fig. 1(b)) and structures which are occluded by the windshield wiper are represented correctly (c.f. Fig. 1(c)).

We evaluate our new approach on a manually labeled driver assistance database of $3,000$ frames with emphasis on adverse weather conditions [9]. It turns out that the number of false positives is reduced significantly, whereas the detection rate of obstacles increases. In comparison to our state of the art stereo techniques [10], [11], the approach

978-1-4799-4337-1/14 $31.00 © 2014 IEEE
DOI 10.1109/CRV.2014.11



(a) (b) (c)

Figure 1. Three examples (a) - (c) of a difficult (rainy) scene at a highway scenario: the first row shows the stereo results using SGM. The distance is color encoded. Red stands for near, green for far away obstacles. The second row shows the results of the Stixel World which is computed by the disparity images. Stixels represent vertical obstacles with a variable height and a fixed width. We use the same color encoding as for SGM. The third row shows the warped results of the *Stix-Fusion* technique into the image: red stands for obstacle areas whereas green represents free space. The advantages of the *Stix-Fusion* technique are clearly shown in the examples (b) and (c): because of the temporal filtering *phantoms* do not occur on the ground surface in contrast to the Stixel-World results (b). Furthermore, structures which are occluded by the windshield wiper or even by spray are represented correctly whereas SGM cannot measure these image parts at this current time step (c).

achieves the best performance on this data set.
By leveraging an efficient NVIDIA CUDA framework, our algorithm runs in real-time in our test vehicle, requiring less than 10 ms per frame.

The remainder of this paper is organized as follows: Section II describes 3D intermediate representations and general occupancy grid mapping techniques. Our approach of the temporal Stixel integration is introduced in Section III: we describe the generation of the input data (Section III-A) and present our general probabilistic framework (Section III-B). Furthermore, the measurement model is described in Section III-C. The realization and the technical details of the general framework are given in Section IV. Section V shows experimental results. Section VI closes this paper with a summary and an outlook.

II. Related Work

The real-time estimation of free space and the detection of obstacles is a fundamental task in robotics. We limit ourselves to related work in techniques which have inspired us for the current approach.

To describe free space and obstacles in a compact fashion, we rely on the Stixel World [8] which is computed from dense disparity images. Pfeiffer et al. [8] formulated a global optimization using a probabilistic framework. The relevant information is represented with a few hundreds Stixels only instead of thousands of stereo depth measurements.

In [12], Benenson et al. described a method to estimate Stixels without computing a disparity map which also reduces the computational costs.

To reconstruct complete 3D environments of urban scenes, Gallup et al. [13] presented a probabilistic method using street-level videos or photo collections. Either structure from motion or dense stereo techniques are used to estimate the required depth maps. A so called n-layer height map is computed which allows the representation of overhanging structures. Based on [13], Zheng et al. [14] presented an efficient incremental depth map fusion framework with the help of wavelet based compression techniques.

Pirker et al. [15] used a Microsoft's Kinect to model the environment in a fast and accurate way using probabilistic occupancy grid mapping techniques [16].

In [17], Badino et al. integrated stereo measurements over time to reduce the disparity uncertainty with the help of a 1D disparity-low-pass filter. Further in [18], the authors mentioned that an integration of Stixels over time would lead to further improvements of free space and obstacle information.

To the best of our knowledge, we picked up the idea for the first time in [19]. 2D Cartesian occupancy grid mapping techniques [20]–[22] were used to achieve an incremental integration of Stixels over time. However, the focus was on building a global map which only includes static environment information. Furthermore, the mapping process was formulated with evidential theory and the map update was realized by the Dempster's Rule of Combination [23]. The sensor model in the approach [19] was defined in the Cartesian 2D space which leads to an inefficient free space and obstacle estimation during the map update. Because of that, the technique is not suitable for real-time applications. In contrast to [19], our present approach is formulated in a Bayesian manner and is based on existence estimation methods. Due to a efficient map representation, the approach runs in real-time in our test vehicle.

III. Probabilistic Framework of the Stixel Integration

Because of the discussion in [24], we formulate the new approach in a probabilistic way. The probability of existence for each grid cell is estimated in a time recursive manner as described below. In this work, an existing cell is occupied and a non-existing cell represents free space. We define a 2D Cartesian fusion map $\mathcal{M} = \{m_{xy}\}$ where m_{xy} represents a grid cell with the coordinates x and y referred to the (global) reference space. This representation is preferable compared to a column-disparity map since the integration of the ego motion and the fusion step can be modeled much easier.

In the following we explain our input data (III-A) and describe the general framework of the recursive existence estimation (III-B). In (III-C), we define our efficient measurement model.

A. Input data

First of all, dense disparity images are estimated using SGM [6] as shown in Fig. 1. In addition, for each disparity value confidence cues are computed [10].

In the next step, the multi-layered Stixel World [8] is computed. Stixels represent the relevant information of the current scene in terms of free space and obstacles (c.f. Fig. 1). A single Stixel is a vertically oriented rectangle with a fixed width w in the image (e.g. $w = 5\,\text{px}$) and a variable height. Referred to the column-disparity-space (u-d-space) each Stixel is defined by a seven dimensional vector $\boldsymbol{s} = [u, v_{top}, v_{base}, w, d, \sigma_d^2, c]$. Here, u is the image column and v_{top} and v_{base} mark the top and base point of the current Stixel in image coordinates. The disparity value of a Stixel is d with its variance σ_d^2. The Stixel confidence information is described by $c \in [0...1]$ which is based on the disparity confidence cues. Further details of the estimation of this cue are described in [10]. All Stixels are collected up to time step t into the matrix structure $\mathcal{S}_{1:t} = [\boldsymbol{s}_1, ..., \boldsymbol{s}_t]$.

For the temporal integration of Stixels, it is essential to know the ego motion of the vehicle. Common techniques are the use of visual odometry [25] or the use of GPS in combination with an inertial measurement unit [19]. In this work, we rely on an inertial measurement unit and define the ego motion between two consecutive time steps $i-1$ and i as the homogeneous motion matrix

$$ {}^{i}\mathbf{X}_{i-1} = {}^{i}\begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \mathbf{0}^T & 1 \end{bmatrix}_{i-1} \tag{1} $$

with the 2D rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{T}$. The complete motion matrix $\mathbf{X}_t$ up to time step t is obtained as the products of the consecutive motion matrices:

$$ \mathbf{X}_t = \prod_{i=1}^{t} {}^{i}\mathbf{X}_{i-1} \ . \tag{2} $$

B. General Framework of the recursive existence estimation

As it is a frequent practice in mapping [16], [20], the global posterior of the map $\mathcal{M}_t$ given the sensor readings $\mathcal{S}_{1:t}$ and the ego motion $\mathbf{X}_t$ is formulated as

$$ p(\mathcal{M}_t|\mathcal{S}_{1:t}, \mathbf{X}_t) = \prod_{xy} p(m_{xy,t}|\mathcal{S}_{1:t}, \mathbf{X}_t) \ , \tag{3} $$

where the grid cells $m_{xy,t}$ are independent from each other.

To formulate $p(m_{xy,t}|\mathcal{S}_{1:t}, \mathbf{X}_t)$ as a time recursive existence estimation problem, the following assumptions are made:

- The sensor readings between consecutive time steps are modeled as conditionally independent and the current cell state $m_{xy,t}$ includes all information from previous sensor readings, resulting in: $p(\boldsymbol{s}_t|m_{xy,t}, \mathcal{S}_{1:t-1}) \approx p(\boldsymbol{s}_t|m_{xy,t})$.
- The state of a grid cell is assumed to be of first-order Markovian nature. This means, the probability of the existence of a grid cell depends only on the direct predecessor: $p(m_{xy,t}|m_{xy,t-1}, ..., m_{xy,0}) \approx p(m_{xy,t}|m_{xy,t-1})$.
- The given ego motion provides no information about the state of a cell: $p(m_{xy,t}|\mathcal{S}_{1:t-1}, \mathbf{X}_t) \approx p(m_{xy,t}|\mathcal{S}_{1:t-1})$.

With theses assumptions it holds, that

$$ p(m_{xy,t}|\mathcal{S}_{1:t}, \mathbf{X}_t) \approx \frac{p(\boldsymbol{s}_t|m_{xy,t}, \mathbf{X}_t)\ p(m_{xy,t}|\mathcal{S}_{1:t-1})}{p(\boldsymbol{s}_t|\mathcal{S}_{1:t-1}, \mathbf{X}_t)} \ . \tag{4} $$

Eq. 4 represents the probabilistic model of our recursive existence estimation problem. It describes the fusion map update over consecutive time steps. Here, $p(\boldsymbol{s}_t|m_{xy,t}, \mathbf{X}_t)$ is the likelihood term which represents the measurement behavior given an existing cell and the ego motion. The definition of this term and further details are described in III-C. The prediction step $p(m_{xy,t}|\mathcal{S}_{1:t-1})$ transforms the state of the cell into the next time step without the actual sensor measurements. Using a Markovian two state transition model the prediction is formulated as:

$$ \begin{aligned} &p(m_{xy,t}|\mathcal{S}_{1:t-1}) = \\ &p(m_{xy,t}|m_{xy,t-1})p(m_{xy,t-1}|\mathcal{S}_{1:t-1}, \mathbf{X}_{t-1}) + \\ &p(m_{xy,t}|\neg m_{xy,t-1})[1 - p(m_{xy,t-1}|\mathcal{S}_{1:t-1}, \mathbf{X}_{t-1})] \ . \end{aligned} \tag{5} $$

The Markov chain with two states is shown in Fig. 2. In contrast to the transition probabilities $p(m_{xy,t}|m_{xy,t-1})$ and $p(\neg m_{xy,t}|\neg m_{xy,t-1})$, the probabilities from one state to the other are usually chosen to be very small (e.g. 0.01). Here, $\neg m_{xy,t}$ means that the cell is not occupied. In [26], a similar proceeding was used to estimate confidence probabilities for vehicle tracking.

Finally, the denominator of Eq. 4 is the normalization term and can be expressed as:

$$p(\boldsymbol{s}_t|\boldsymbol{S}_{1:t-1}, \mathbf{X}_t) = \\ p(\boldsymbol{s}_t|m_{xy,t}, \mathbf{X}_t)p(m_{xy,t}|\boldsymbol{S}_{1:t-1}) \; + \\ p(\boldsymbol{s}_t|\neg m_{xy,t}, \mathbf{X}_t)[1 - p(m_{xy,t}|\boldsymbol{S}_{1:t-1})] \; . \tag{6}$$

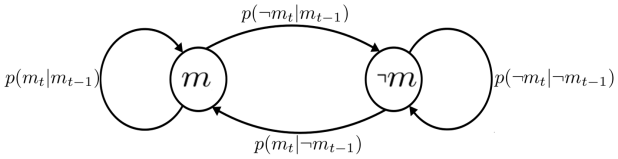


Figure 2. The Markovian two state transition model. It is used for the prediction step during our recursive existence estimation. For temporal stability the transition probabilities from one state to the other are usually chosen much smaller than transitions back into the same state.

C. Definition of an efficient measurement model

In order to define the measurement model, one challenge exists: The current Stixel $\boldsymbol{s}_t$ is described in the column-disparity-space (u-d-space) whereas the fusion map $\mathcal{M}_t$ is characterized in a Cartesian way. Following Eq. 4, a map update in the Cartesian space is required. For an efficient definition of our measurement model, we use a *from-target-to-source*-strategy where the *target* represents the Cartesian fusion map and a Stixel measurement represents the *source*.

First, we transform the coordinates of a cell $[x \; y \; 1]^T$ into local coordinates $[x' \; y' \; 1]^T$ which are referred to the ego vehicle coordinate system:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = {\mathbf{X}_t}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} . \tag{7}$$

Assuming a pinhole camera model and a stable stereo camera calibration during the acquisition time, the projection into the u^*-d^*-space is

$$d^* = \frac{f \; b}{x'} \quad \text{and} \quad u^* = u_0 + \frac{f \; y'}{x'}. \tag{8}$$

Here, f is the focal length and u_0 the column part of the principal point of the left camera. b represents the base line of the stereo rig. With the help of the calculated disparity and column values (d^* and u^*) for a given grid cell m_{xy}, the measurement model is defined as a mixture model composed of a Gaussian and a uniform distribution:

$$p(\boldsymbol{s}_t|m_{xy,t}, \mathbf{X}_t) = \\ \alpha \; p_{norm}(d_t|d^*, \sigma_{d^*}^2) + (1 - \alpha) \; p_{uni}(d_t|d^*, \sigma_{d^*}^2) \; . \tag{9}$$

Similar to the described forward model in [27], we define the Gaussian distribution as

$$p_{norm}(d_t|d^*, \sigma_{d^*}^2) = \\ \begin{cases} \eta \mathcal{N}(d_t|d^*, \sigma_{d^*}^2) & \text{if } d_{max} > d_t > d_{min} \text{ and} \\ & \quad u^* = u_t \\ 0 & \text{else} \end{cases} \tag{10}$$

and the uniform distribution as

$$p_{uni}(d_t|d^*, \sigma_{d^*}^2) = \\ \begin{cases} \frac{1}{(d_{max} - d_{min})} & \text{if } d_{max} > d_t > d_{min} \text{ and} \\ & \quad u^* = u_t \\ 0 & \text{else} \; . \end{cases} \tag{11}$$

Eq. 10 and Eq. 11 represent an efficient measurement model because it takes into account only the disparity value of the current Stixel d_t. In contrast to [4], an uncertainty in the disparity space is only assumed.

The measurement model given non-existing cells is formulated as

$$p(\boldsymbol{s}_t|\neg m_{xy,t}, \mathbf{X}_t) = \\ \alpha \; \widetilde{p}_{norm}(d_t|d^*, \sigma_{d^*}^2) + (1 - \alpha) \; p_{uni}(d_t|d^*, \sigma_{d^*}^2) \tag{12}$$

with

$$\widetilde{p}_{norm}(d_t|d^*, \sigma_{d^*}^2) = \\ \begin{cases} \widetilde{\eta}\big(1 - \mathcal{N}(d_t|d^*, \sigma_{d^*}^2)\big) & \text{if } d_{max} > d_t > d_{min} \text{ and} \\ & \quad u^* = u_t \\ 0 & \text{else} \; . \end{cases} \tag{13}$$

Here, η and $\widetilde{\eta}$ are the normalizers. The *max*- and *min*-values of the disparity range are described by d_{max} and d_{min}. The variance $\sigma_{d^*}^2$ and the weighting factor α are the model parameters and have to be learned iteratively [16].

IV. Realization of the Framework and Technical Details

In this work, the general framework of the Stixel integration is used to estimate robust free-space and obstacle information. The efficient realization of our concept is shown in Fig. 3.

Since we define the disparity range by $d^* \in [0...128]$, we start at the *source* directly. With the defined measurement model a local u^*-d^*-probability map (c.f. Fig. 3, image 1) is computed similar to [15], [17], [28] . Due to the closed interval of the disparity space, the normalizers η and $\widetilde{\eta}$ are known.

In a simplified manner, the model parameter $\sigma_{d^*}^2$ is set to the disparity variance of each Stixel $\sigma_{d_t}^2$. The weighting factor α is defined by the confidence value of a Stixel c_t.

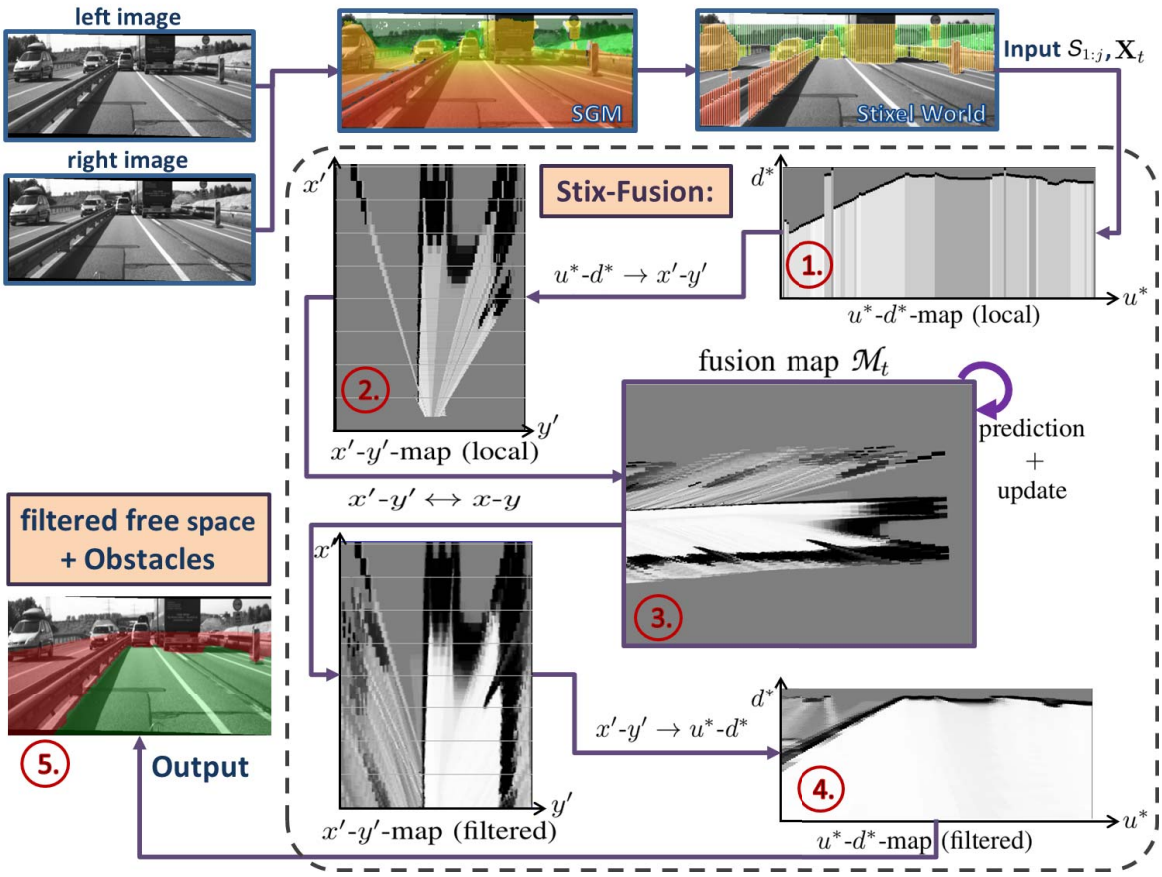


Figure 3. The realization of the general framework (highlighted by the dashed line): with the help of the left and right image sequences dense disparity images are estimated via SGM. The Stixel World is computed and is used as input data. A column (u^*)-disparity (d^*)-probability map is computed for each time step which is shown in image 1. The transformation into the local x'-y'-space is straightforward (image 2). The prediction and update step of the fusion map $\mathcal{M}_t$ is carried out on the GPU. The image 3 illustrates the filter result after 162 time steps. A back transformation of the filtered map into the u^*-d^*-space is shown in image 4. The results are warped into the current image frame (image 5) where green represents free space and red obstacles.

Subsequently, the transformation into the x'-y'-space follows (c.f. Fig. 3, image 2). For the prediction step we have to define the two transition probabilities with $p(m_{xy,t}|m_{xy,t-1}) = 0.99$ and $p(\neg m_{xy,t}|m_{xy,t-1}) = 0.01$. As a consequence, our system achieves a strong low pass filter effect. Therefore, a reduction of outliers can be expected. A drawback of this setting is that dynamic obstacles produce a confidence trail (c.f. Fig. 3, image 3).

Since we are only interested in the surroundings of our ego vehicle, we limit the fusion map $\mathcal{M}_t$ to $100 \times 100\,m^2$ with a cell grid resolution of 0.1 m. Note that the general framework can be used to build global 2D occupancy grid maps as long as the ego motion is known.

After the fusion and prediction step, an inverse transformation is carried out to represent a filtered probability map in the column-disparity-space (Fig. 3, image 4). Finally, this map is warped into the image plane for a comprehensible representation of free space and obstacles (c.f. Fig. 3, image 5). All relevant parameters are collected in Tab. I. Just like SGM [7], the Stixel World computation is performed on a FPGA platform at 25 Hz. The presented approach is running using the GPU (NVIDIA GeForce GTX 480) and CPU (Intel Core i7-980X 3.33Ghz) in under 10 ms in our vehicle.

Table I
SETTINGS OF MOST RELEVANT PARAMETERS:

parameter	definition
d^*	$\in [0...128]$
$\sigma^2_{d^*}$	$\sigma^2_{d_t}$
α	c_t
$p(m_{xy,t}\|m_{xy,t-1})$	0.99
$p(\neg m_{xy,t}\|m_{xy,t-1})$	0.01
size of $\mathcal{M}_t$	$100 \times 100\,m^2$
grid cell resolution	$0.1\,m$

V. Evaluation

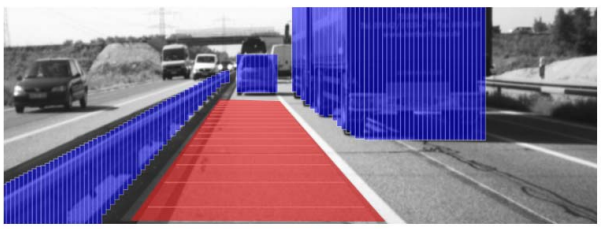
Our presented technique is evaluated on a manually labeled dataset which is publicly available [9]. It consists of 12 sequences with a total number of about $3,000$ frames. The database depicts mostly rainy highway scenes with blurred windshield, wiper passes, and spray water. The dataset describes all ground truth 3D obstacles limiting the available free space referring to the ego vehicle (c.f. Fig. 4).

To evaluate our *Stix-Fusion* technique on this dataset we have to classify our probabilistic maps into free space and obstacle information. To this end, the following steps are carried out:

1) Binarization of the the filtered u^*-d^*-probability map into free and occupied cells. The threshold for occupied cells is empirically set to 0.85.
2) Extraction of the nearest obstacle for each column. Up to the first obstacle we define free space.
3) Transformation of each obstacle into the 3D Cartesian space of the ego vehicle under the assumption of a fixed height (e.g. $1.5\,m$).

After these steps it is possible to compute the detection rate as well as the false positive rate referring to the ground truth dataset. False positives are defined by obstacles which occupy the free space area. The detection rate is estimated by counting all obstacles which are conform with the manually labeled structures. We compare our results to the following three techniques:

- Baseline (*Stix*): Results of the Stixel World approach [8] using SGM stereo [6] without confidence estimation.
- Stixels with confidences (*StixConf*): Results of the Stixel World using SGM stereo with confidence estimation as described in [10].
- Stixels with temporal stereo scene priors (*StixScenePrior*): Results of the Stixel World using stereo which is based on a graph-cut technique with temporal scene priors [11]. Note that no stereo confidences (and therefore no Stixel confidences) are available.



(a)

Figure 4. An example of the manually labeled ground truth. Blue regions represent obstacles whereas the predicted driving corridor for the ego vehicle is red.

The input for the *Stix-Fusion* are the Stixels with confidences as described in Section III-A. During the evaluation we only vary the transition probability $p(m_{xy,t}|m_{xy,t-1})$ because it is the key parameter of our framework. As an example, *StixConfFusion90* stands for a transition probability of 0.90. All other parameters were fixed (c.f.Tab. I).

The results of our evaluation are shown in Tab. II. We obtain 200 frames containing false positives in the baseline approach (*Stix*). When using *Stix-Fusion*, the number of frames with false positives is 31 by *StixConfFusion90*, 32 by *StixConfFusion95* and 22 by *StixConfFusion99*.

Compared to the Stixel results with confidences, the number of frames with false positives is reduced by a factor of two (*StixConf* with 45 frames with fp vs. 22 frames with fp). The total number of false positives is reduced by a total number of 45 (107 fp vs. 62 fp). Note that the detection rate increases to 88.2 % whereas *StixConf* achieved only a rate of 80.5 %.

In a further step we use the baseline as input data and set the transition probability to 0.99 (*StixFusion99*). This affects our method in two ways: our algorithm has to handle about seven times more false positives (107 fp vs. 754 fp) as in the experiments above. Furthermore, the weighting factor α is static and is defined by 0.5.

The comparison of *StixConfFusion99* and *StixFusion99* underlines the robustness of the approach against outliers. In spite of a seven times higher number of outliers in the input data, the modification achieves only 15 more false positives (62 fp vs. 77 fp) whereas the detection rate does not change significantly (88.2 % vs. 88.5 %). The detection rate of 88.5 % was only obtained by the *StixScenePrior* approach, but the number of false positives is nearly a factor of five higher. In addition, one can clearly see that our approach is not very sensitive in the model parameter α. With respect to the temporal integration this value becomes less important.

The evaluation shows that *Stix-Fusion* improves the estimation of free space and obstacles even in challenging scenarios. Fig. 5 shows example results which underline this statement. In comparison to the Stixel results no *phantoms* appear in free space areas and the guard rails and vehicles are detected correctly.

Table II
Comparison of false positive stixels (fp), number of frames with false positives (fp frames), and detection rates on the Ground Truth Stixel Database

approach	Number of fp	frames with fp	detection rate
Stix	754	200	81.5 %
StixConf	107	45	80.2 %
StixScenePrior	658	141	**88.5** %
StixConfFusion90	66	31	81.6 %
StixConfFusion95	65	32	85.5 %
StixConfFusion99	**62**	**22**	88.2 %
StixFusion99	77	30	**88.5** %

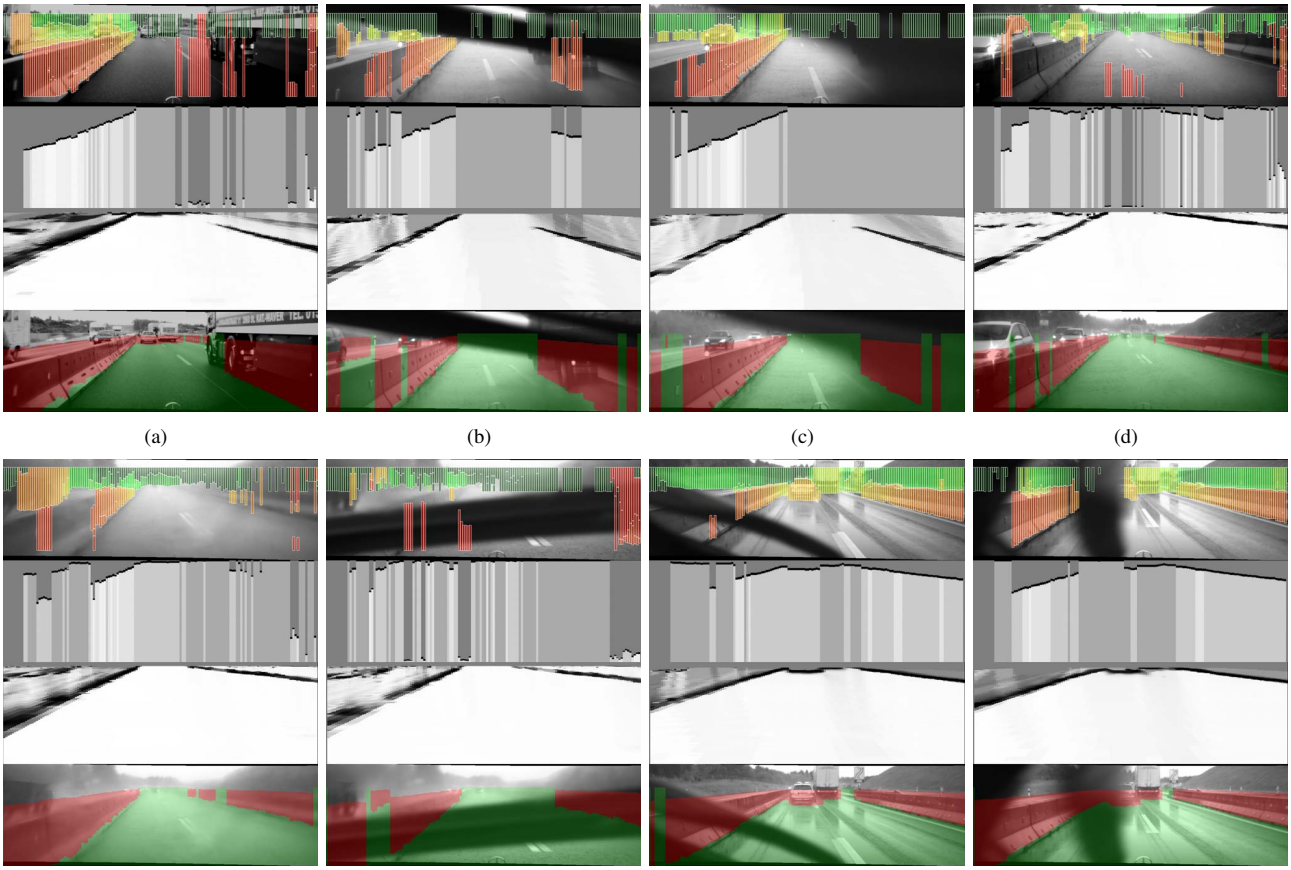


Figure 5. Results of our *Stix-Fusion* technique during difficult weather conditions on a highway. All examples (a)-(h) have the same order: the first image shows the Stixel input data [8], the second image represents the local and the third the filtered u^*-d^*-probability map. Under the assumption of a fixed height and a binarized map, these results are warped into the image (fourth image). Here, red stands for obstacles and green for free space. In spite of the partly moderate Stixel results caused by the windshield wiper, even in these scenes the guard rails and vehicles are reconstruct correctly by the *Stix-Fusion* process. Furthermore, false Stixel measurements (*phantoms*) do not occur on the ground surface.

VI. Summary and Outlook

In this work, we presented a robust estimation of free space and occupied areas using stereo vision in combination with probabilistic occupancy grid mapping techniques. Therefore, we relied on the so called Stixel World, a super pixel representation which is based on disparity images. By means of a recursive existence estimation, Stixels are fused into a reference grid map in a probabilistic fashion.

We show that our *Stix-Fusion* technique achieved the best results on a manually labeled $3,000$ image dataset compared with other stereo approaches. The dataset contains 3D ground truth obstacle data as well as free space information which is publicly available.

Furthermore, we have shown that the new approach can handle a high rate of outliers whereas the detection rate is at a near constant level. Even under adverse weather conditions our algorithm estimates free space and obstacles in a reliable manner.

In future, our goal is to handle dynamic obstacles because up to this point they produce undesirable trails in our filtered maps. As described in [29], Stixel tracking is a proven method which will help to solve this challenge in future.

In addition, we want to use the height information of the Stixels to produce a height layer similar to [4]. It is to be expected that these improvements reduce the false positive rate again and increase the detection rate as well.

References

[1] S. Thrun, "What we're driving at," http://googleblog.blogspot.de/2010/10/what-were-driving-at.html, October 2010.

[2] S. Kammel, J. Ziegler, B. Pitzer, M. Werling, T. Gindele, D. Jagzent, J. Schröder, M. Thuy, M. Goebl, F. v. Hundelshausen, O. Pink, C. Frese, and C. Stiller, "Team annieway's autonomous system for the 2007 DARPA Urban Challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 615–639, 2008.

[3] U. Franke, D. Pfeiffer, C. Rabe, C. Knöppel, M. Enzweiler, F. Stein, and R. G. Herrtwich, "Making bertha see," in *ICCV Workshop on Computer Vision for Autonomous Driving*, Sydney, Australia, September 2013.

[4] M. Muffert, D. Pfeiffer, and U. Franke, "A stereo-vision based object tracking approach at roundabouts," in *IEEE Intelligent Transportation Systems Magazine*, Volume 5, Number 2, Summer 2013, pp. 22–23.

[5] "VisLab PROUD-Car Test 2013 Webpage," http://www.vislab.it/proud/, August 2013.

[6] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, June 2005, pp. 807–814.

[7] S. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using semi-global matching," in *International Conference on Computer Vision Systems (ICVS)*, 2009.

[8] D. Pfeiffer and U. Franke, "Towards a global optimal multi-layer Stixel representation of dense 3D data," in *British Machine Vision Conference BMVC*. Dundee, Scotland: BMVA Press, August 2011.

[9] http://www.6d-vision.com/ground-truth-stixel-dataset.

[10] D. Pfeiffer, S. Gehrig, and N. Schneider, "Exploiting the power of stereo confidences," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, USA, June 2013.

[11] S. Gehrig, M. Reznitskii, N. Schneider, and U. Franke, "Priors for stereo vision under adverse weather conditions," in *ICCV Workshop on Computer Vision for Autonomous Driving*, Sydney, Australia, September 2013.

[12] R. Benenson, R. Timofte, and L. Van Gool, "Stixels estimation without depth map computation," in 2^{nd} *IEEE Workshop on Computer Vision in Vehicle Technology: From Earth to Mars (CVVT) in conjunction with the* 13^{th} *International Conference on Computer Vision (ICCV))*, November 2011.

[13] D. Gallup, M. Pollefeys, and J.-M. Frahm, "3d reconstruction using an n-layer heightmap," in *German Association for Pattern Recognition (DAGM)*, Darmstadt, Germany, September 2010, pp. 1–10.

[14] E. Zheng, E. Dunn, R. Raguram, and J.-M. Frahm, "Efficient and scalable depthmap fusion," in *British Machine Vision Conference (BMVC)*, Surrey, England, December 2012, pp. 34.1 –34.12.

[15] K. Pirker, G. Schweighofer, M. Ruether, and H. Bischof, "Fast and accurate environment modeling using three-dimensional occupancy grids," in *Proc. 1st IEEE Workshop on Consumer Depth Camera for Computer Vision (ICCV,CDC4CV)*, 11 2011.

[16] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, ser. Intelligent Robotics and Autonomous Agents. The MIT Press, 2005.

[17] H. Badino, U. Franke, and R. Mester, "Free space computation using stochastic occupancy grids and dynamic programming," in *Workshop on Dynamical Vision, ICCV*, Rio de Janeiro, Brazil, October 2007.

[18] H. Badino, U. Franke, and D. Pfeiffer, "The Stixel World - A compact medium level representation of the 3D-world," in *German Association for Pattern Recognition (DAGM)*, Jena, Germany, September 2009, pp. 51–60.

[19] M. Muffert, S. Anzt, and U. Franke, "An incremental map building approach via static stixel integration," in *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., II-3/W3, 55-60, doi:10.5194/isprsannals-II-3-W3-55-2013*, Antalya, Turkey, September 2013.

[20] A. E. Elfes, "Sonar-based real-world mapping and navigation," *Journal of Robotics and Automation*, vol. 3, no. 3, pp. 249–265, June 1987.

[21] D. Murray and J. J. Little, "Using real-time stereo vision for mobile robot navigation," *Autonomous Robots*, vol. 8, no. 2, pp. 161–171, 2000.

[22] J. Moras, V. Cherfaoui, and P. Bonnifait, "Moving objects detection by conflict analysis in evidential grids," in *IEEE Intelligent Vehicles Symposium (IV)*, Baden-Baden, Germany, June 2011, pp. 1122–1127.

[23] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[24] P. Cheeseman, "In defense of probability," in *In Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 1985, pp. 1002–1009.

[25] H. Badino, A. Yamamoto, and T. Kanade, "Visual odometry by multi-frame feature integration," in *ICCV International Workshop on Computer Vision for Autonomous Driving*, Sydney, Australia, December 2013.

[26] R. Altendorfer and S. Matzka, "A confidence measure for vehicle tracking based on a generalization of bayes estimation," in *IEEE Intelligent Vehicles Symposium (IV)*, San Diego, CA, USA, 2010.

[27] S. Thrun, "Learning occupancy grid maps with forward sensor models," *Autonomous Robots*, vol. 15, pp. 111–127, September 2003. [Online]. Available: http://portal.acm.org/citation.cfm?id=940152.940193

[28] M. Perrollaz, J.-D. Yoder, A. Spalanzani, and C. Laugier, "Using the disparity space to compute occupancy grids from stereo-vision." in *IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*. IEEE, 2010, pp. 2721–2726.

[29] D. Pfeiffer and U. Franke, "Efficient representation of traffic scenes by means of dynamic Stixels," in *IEEE Intelligent Vehicles Symposium (IV)*, San Diego, CA, USA, June 2010, pp. 217–224.