

## Exploiting Traffic Scene Disparity Statistics for Stereo Vision

Stefan K. Gehrig  
Daimler AG  
HPC 050-G024  
71059 Sindelfingen, Germany

Uwe Franke  
Daimler AG  
HPC 050-G024  
71059 Sindelfingen, Germany

Nicolai Schneider  
IT-Designers GmbH  
Entennest 23  
73070 Esslingen, Germany

### Abstract

*Advanced Driver Assistance Systems benefit from a full 3D reconstruction of the environment in real-time, often obtained via stereo vision. Semi-Global Matching (SGM) is a popular stereo algorithm for solving this task which is already in use for production vehicles. Despite this progress, one key challenge remains: stereo vision during adverse weather conditions such as rain, snow and low-lighting.*

*Current methods generate many disparity outliers and false positives on a segmentation level under such conditions. These shortcomings are alleviated by integrating prior scene knowledge. We formulate a scene prior that exploits knowledge of a representative traffic scene, which we apply to SGM and Graph Cut based disparity estimation. The prior is learned from traffic scene statistics extracted during good weather. Using this prior, the object detection rate is maintained on a driver assistance database of 3000 frames including bad weather while reducing the false positive rate significantly. Similar results are obtained for the KITTI dataset, maintaining excellent performance in good weather conditions.*

*We also show that this scene prior is easy and efficient to implement both on CPU platforms and on reconfigurable hardware platforms. The concept can be extended to other application areas such as indoor robotics, when prior information of the disparity distribution is gathered.*

### 1. Introduction

Stereo vision has been an active area of research for decades. Recent years have shown a trend towards global stereo algorithms that optimize the disparity map jointly, rather than individually for each pixel [1]. The Middlebury database [1] is a good resource of available stereo algorithms, but its scene complexity is limited. A more challenging benchmark is the KITTI database [2], comprising of some 200 image pairs of street scenes. It still underrepresents the challenges for vision-based advanced driver assistance systems that should operate at all weather and

illumination conditions, such as rain, snow, night, and combinations thereof. These challenging scenarios inspired our work, in which we reduce the disparity errors by introducing prior knowledge into the estimation process.

In the light of increasing autonomy of future vehicles, such adverse weather scenarios have to be mastered. Work on benchmarking such scenarios has just recently started. The Heidelberg HCI dataset [3] was the first data set covering challenging weather scenarios, however, without supplying ground truth. The Ground Truth Stixel Dataset [4] contains a set of rainy highway scenes with sparse ground truth labels for the free space and objects.

For driver assistance, the immediate surroundings of the car that limit the free space should be detected at all times but without mistakenly detecting a structure within the free space. An successful example for solving this task in real-time is Semi-Global Matching [5] (SGM), which can also be found in the top 10 of the KITTI benchmark.

Under adverse weather conditions, SGM has a uniform disparity distribution for outlier disparities. Mapping the distribution into 3D space, we measure most outliers right in front of the car. To counteract this observation we introduce a scene prior: Using statistics drawn from many traffic scenes we generate a representative traffic scene and use this information as a prior for the disparity estimation process. This generates a small bias towards the representative traffic scene when basically no other data is available from the image pair, e. g. in regions occluded by the windshield wiper. An example with windshield wiper is shown in Figure 1.

We introduce this new scene prior and apply it both to SGM and Graph Cut Stereo (GC) resulting in a significant reduction of false positives under adverse weather conditions. Summarizing, the main contributions of this paper are: the generation of a representative traffic scene, the introduction of an efficient and effective scene prior, applicable to many stereo algorithms; and a careful evaluation of the new algorithm variants on KITTI data and on a 3000-frames highway database with manually labeled ground truth that includes adverse weather conditions.

The rest of the paper is organized as follows. Section 2



Figure 1. Rain traffic scene. Stereo reconstruction (red=near ... green=far) for the scene using SGM (left), right image (center), and SGM with the scene prior introduced here (right). Large red blobs indicate nearby objects leading to potential false positive objects.

covers related work in stereo to incorporate priors. Section 3 describes our used stereo methods, Graph Cut and SGM, briefly. In Section 4 we detail how to incorporate the scene prior in a Bayesian framework. Implementation details for CPU and FPGA implementation are discussed in Section 5. Section 6 shows results for Graph Cut, SGM and the new prior on a 3000-frames database with challenging highway driving scenes. With this, false positive point statistics and detection rates are presented on pixel and intermediate level for the new stereo variant introduced in this paper.

## 2. Related Work

We limit ourselves to related work in stereo vision using priors. "Prior" in the context used here means prior information that is independent of the image information in the current stereo pair. Related work on stereo hardware implementations is listed in Section 5.2.

One popular option is to use smoothness priors. Standard smoothness priors assume a piece-wise constant depth in the scene (e. g. [6]). Other works try to enforce a second-order smoothness prior that allows for reconstructing slanted surfaces, e. g. [7]. This has been tried with moderate success with SGM as well [8]. Another smoothness prior enforces a disparity ordering along columns due to the scene layout in traffic scenes for SGM[9].

Shape priors are popular in multi-view stereo (see e. g. [10]). Planarity and orthogonality have been exploited as priors several times, e. g. in [11].

Scene prior in our context means exploiting prior knowledge about the scene from other sources than the image pair itself. In [12], a sparse point cloud obtained from structure-from-motion is used as a scene prior to render reconstructions deviating from the sparse result less likely. This information is injected by modifying the data term for the respective points.

The method closest to our work is the scene prior from [13]. There, a simple scene prior that renders larger disparities less likely in general is used. All pixels in the image are treated the same way which inhibits good reconstructions of the nearby road profile since the prior is heavily violated. We compare our results to that method in Section 6. Scene priors using information about a representative traffic scene have not been used before to the best of our knowledge.

## 3. Semi-Global Matching and Graph Cut

### 3.1. Semi-Global Matching

Roughly speaking, SGM [5] performs an energy minimization on multiple independent 1D paths crossing each pixel and thus approximates a 2D connectivity. After cost accumulation the classic winner-takes-all approach is applied. The energy consists of three parts: a data term for similarity, a small penalty term for slanted surfaces that change the disparity slightly (parameter  $P_1$ ), and a large penalty smoothness term for depth discontinuities ( $P_2$ ).

Hirschmueller *et al.* [14] achieved very good performance SGM using the Hamming distance of images transformed with a 9x7 Census as similarity criterion [15]. Other investigations (e. g. [16], [17]) have confirmed this finding. We apply this similarity criterion throughout our work.

In order to identify and exclude occlusions, Hirschmueller [5] performs one SGM run with the left image as the reference image and another run with the right image as the reference image. Disparities that deviate by more than 1 pixel between the two results are removed (RL-consistency check).

### 3.2. Graph Cut

Graph Cut (GC) is the method of choice for an efficient near-optimal solution to multi-label problems. The first known GC stereo algorithm was introduced by Boykov *et al.* [6]. For our investigations, we use the implementation from [18] without label costs. There are two variants of GC: alpha-expansion and swap-label. The first algorithm requires a metric, while the second can also operate on semi-metrics. The parallels between GC and SGM have been shown in [13]. There, it was shown that SGM and GC performed very similarly when using the same smoothness potential and the same similarity criterion, namely the hamming distance of a 9x7 Census. The improvements applying priors in addition lead to almost identical results for both methods. We use the same parametrization as in [13] for our evaluations.

For reference, we run GC and SGM with the same parameters, the same data term, and the same right-left check mechanism. Sub-pixel interpolation is not performed since robustness, not accuracy, is our main concern. On good weather data, the two variants exhibit very little difference.

An example is shown in Figure 2. We perform all downstream evaluations with GC as well as with SGM in order to show the effectiveness of the prior being independent of the stereo method.



Figure 2. Standard traffic scene overlaid with disparity result SGM (left) and GC (right). Red pixels are near, green far away.

## 4. Scene Prior

In adverse weather, a stereo algorithm that relies solely on the image data has intrinsic performance limits due to image noise or disturbances of the optical path as shown in Figure 1. An additional prior is able to stabilize the disparity maps in such situations.

### 4.1. Incorporation of the Scene Prior

To show how the prior can be incorporated we describe our stereo task in a probabilistic fashion and extend it with the new scene prior. We seek the disparity map  $D$  that maximizes the probability

$$p(D|I_L, I_R) \propto p(I_L, I_R|D) \cdot p(D), \quad (1)$$

where  $I_L/I_R$  is the left/right image,  $p(D)$  represents the prior term. We assume the binary smoothness term to be independent of the unary scene prior term learned from statistics, so we can multiply both terms:

$$p(D) \propto p_{smooth}(D) \cdot p_{val}(D). \quad (2)$$

The first term is the well-known smoothness term and the second one the new scene prior term. The prior term is detailed in Section 4.3. All above terms are carefully normalized to one to obtain their relative weights automatically, without parameter tuning of weights. Before describing the scene prior in more detail, we turn to the offline process of gathering scene prior information.

### 4.2. Obtaining Scene Prior Information

We base our scene prior on the 3D shape of typical traffic scenes. In order to collect data for the representative traffic scene we took a stereo camera system in a vehicle and recorded more than 100000 frames under good weather conditions. The driving mix contains mostly urban and rural roads. We collect all estimated disparities for every pixel in a histogram and normalize them to one (see Figure 4). This data is collected with a tilt estimation module active to compensate tilt motion of the vehicle [19]. From these

histograms, we obtain the most probable disparity per pixel at the peak of the histogram. We call this disparity the average disparity for brevity knowing that it is not equivalent to the average disparity. The resulting average disparity image  $\bar{D}$  for our stereo camera system is shown in Figure 3 top. Below the associated occurrence probability image for the average disparity image is shown. This occurrence probability depicts the probability value of the maximum in the disparity histogram. One can see the street up to the horizon and the sky in the center top of the image.

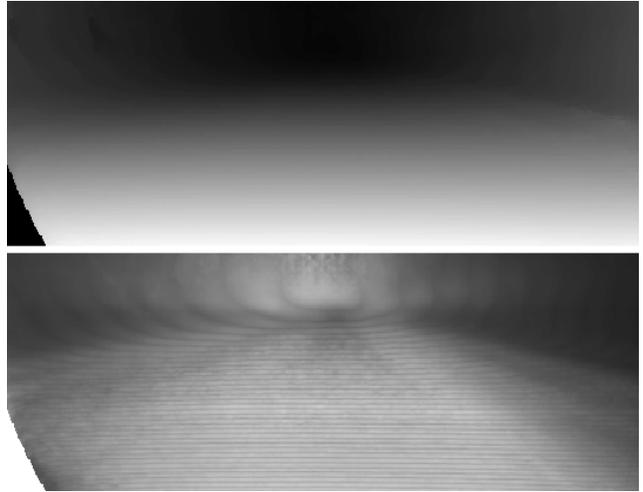


Figure 3. Average disparity map for the representative traffic scene (top). Darker pixels correspond to smaller disparities/larger distances. Occurrence probability map for the representative traffic scene (bottom), the brighter the more likely the respective disparity value occurs.

Some example disparity histograms of the traffic scene statistics are shown in Figure 4. On top, the disparity distribution for a street pixel right in front of the ego-vehicle is shown. We expect to see the street disparity at more or less the same distance. A clearly peaked disparity distribution is obtained, showing nearly 30% probability for the average disparity for the street region (marked in red) right in front of the vehicle. Very few other disparities occur besides the strong peak around disparity value 23. A different distribution is obtained at the far right of the image slightly below the horizon. This distribution is shown in the bottom histogram. Here we expect both small disparities when no obstacles are present and many other possible disparities for obstacles at varying distances. This is confirmed in Figure 4 bottom where many possible disparities occur. Consequently, the resulting most probable peak is much lower with only 3% probability for the most probable disparity value 18 and the distribution is less distinct.

Our stereo camera system is mounted in the car behind the rear-view mirror at about 1.3m height above the street. Other stereo configurations mounted at this height

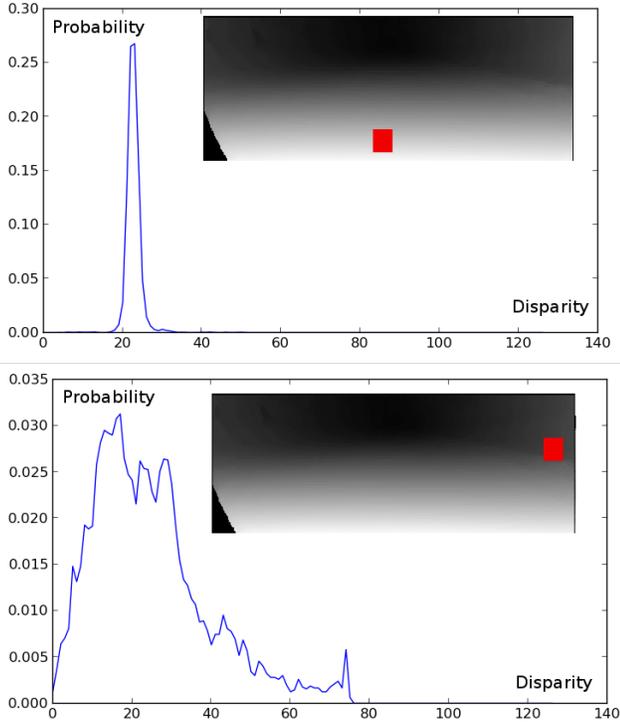


Figure 4. Example histograms of a street pixel close to the vehicle (top) and of a pixel on the far right (bottom) with a much more dispersed distribution.

are mapped onto the same disparity image by normalizing disparities via the product of focal length and baselines and correcting for installation tilt, roll, and yaw angle. This normalization procedure allows for collecting statistics from different camera setups and vehicles. Moreover, one can apply the same disparity statistics to any stereo camera system mounted at that height. In fact, we generated our statistics with a 32cm stereo baseline system and applied it on a data set with 21cm stereo baseline.

We gathered the same statistics for the KITTI stereo data set as well. Here we used all 20 image pairs of the 194 training and testing sequences resulting in about 8000 images. The camera height is significantly above 1.3m so the data cannot be merged. The average disparity image is shown in Figure 5 top. The image below shows the maximum probability image ( $p_{max}$ ) followed by the reciprocal maximum probability image,  $\sigma = \frac{1}{\sqrt{2*\pi*p_{max}}}$ . This value represents the standard deviation when a Gaussian distribution is on hand. At the bottom, the empirical variance  $\sigma_{di}$  computed from the histograms is shown. The two bottommost images show similar values and result in very similar performance when used as prior. We show results using the empirical variance image in the prior computation since this does not imply restrictions on the type of distribution.

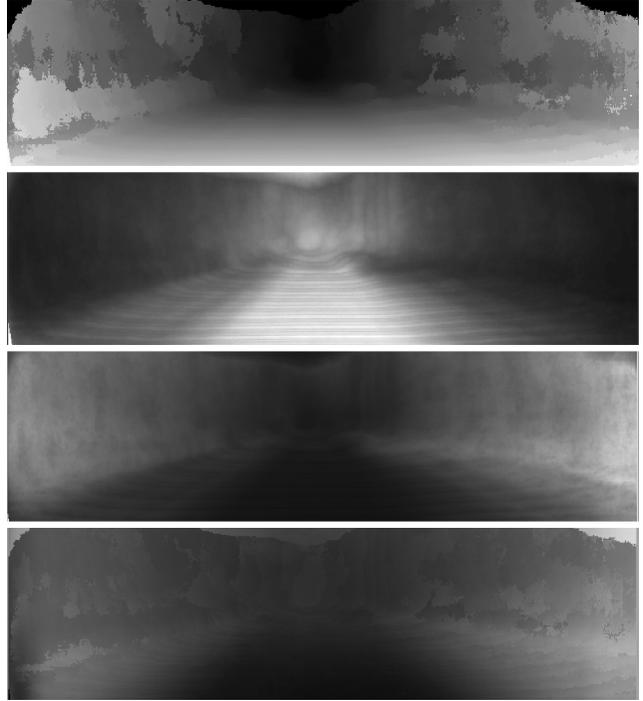


Figure 5. Average disparity map of the KITTI stereo data set (top) and occurrence probability of it underneath. The estimated Gaussian standard deviation and the empirical variance of the disparity distribution are shown in the two bottom rows. For all images holds: the whiter the pixel the higher its value.

### 4.3. Computation of the Scene Prior

Having collected normalized histograms for every pixel in the image, we theoretically could take these histograms as probabilities for every possible disparity per pixel and implement a scene prior from that. However, this would require to read tens of millions of data entries at startup and would lead to zero probability for disparities that did not occur during data acquisition.

Instead, we model the disparity probabilities as a Gaussian distribution around the most probable disparity with a standard deviation determined from the empirical variance  $\sigma_{di}$ . To allow for disparity values that are not reflected in the gathered statistics, we add an additional uniform distribution to the Gaussian distribution to allow for rare disparities and to tune the scene prior effect.

The assumption of the prior information being independent for every pixel leads to a simple addition to the data cost volume, a unary term. This is very efficient to implement. We obtain:

$$p_{val}(D) = \prod_i p_{val}(d_i), \quad (3)$$

$$p_{val}(d_i) = (1 - p_{out})\mathcal{N}(\bar{d}_i, \sigma_{di}) + p_{out} \mathcal{U},$$

where  $\mathcal{N}$  is the normal distribution of  $d_i$  with mean  $\bar{d}_i$  and

standard deviation  $\sigma_{d_i}$  as parameters drawn from the scene statistics ( $\mathcal{N} \propto e^{-(d_i - \bar{d}_i)^2 / 2\sigma_{d_i}^2}$ ),  $\mathcal{U}$  the uniform distribution, and  $p_{out}$  is the outlier probability for the prior information to be wrong. This is the only parameter to explore.

Roughly speaking, we introduce a slight bias towards the average disparity into the data term to prefer disparity values compliant with the gathered statistics. All probabilities introduced above are transferred to log-likelihood energies [20] and these energies are fed into the GC engine from [18]. The priors are easily transferred back to SGM and GC since only the data term is affected.

## 5. Implementation of the Scene Prior

### 5.1. Software Implementation

The introduced scene prior is independent of the chosen stereo method as it only operates on the data term. Looking at Equations 3, we choose not to compute the individual costs per pixel and disparity hypothesis on-the-fly since this would lead to a tremendous computational overhead. These additional scene prior costs can easily be precomputed once at startup since all data is known a priori. An average disparity image and a variance image is read in and Equation 3 is computed for every pixel in the image and for every disparity hypothesis. The results are stored in memory of size 512 (width) · 220 (height) · 64 (disparities) = 7208960 bytes for images downscaled by a factor of 2.

The computational overhead to compute the scene prior after startup is small. With little optimization, the runtime increases by 20ms for the scene prior due to a modular software design. Timings are for a Core(TM) i7 PC on images downscaled by a factor of two, resulting in 512x220px size.

### 5.2. Firmware Implementation for an FPGA

For an efficient hardware implementation on a field-programmable gate array (FPGA) we pick SGM as proposed by Gehrig et al. [21]. Given this basis, we only replace the ZSAD cost generation block shown in Fig. 2 from [21]. We use a 9x7 Census instead of ZSAD, reducing the cost data width from 14 to 6 bit. This Census cost metric is very efficient to implement on reconfigurable hardware (see e. g. [22]). The Census data cost block is extended with the additional data costs from the scene prior (see Figure 6).

Similar to the software implementation, all scene prior calculations are computed once in software at startup. However, it is not trivial to read the necessary statistics data from RAM and add it to the Census data term without additional latency, not the mention the high bandwidth necessary.

Therefore, we decide to store the scene prior costs in the FPGA internal block RAM (BRAM) memory. The net memory size for this is 7.2 million times 4bits for the costs. This part can hardly be handled inside the FPGA since one would need too much BRAM for this informa-

tion which is only available in high-end FPGAs. To keep BRAM consumption acceptably low, we discretize the average disparity values and the disparity standard deviations in integer values, leaving 64 average disparity values and 16 standard deviation values. Only less than 0.5% of the pixels have standard deviation beyond 16 pixels so we loose very little information here. This discretization leaves us with 1024 possible cost vectors. With 4bits maximum prior cost which covers outlier probabilities down to 0.2, we need about  $1024 \cdot 4 \cdot 64\text{bits} = 256\text{kbits}$  to store all possible additional cost vectors for a pixel. This memory demand can easily be handled by automotive FPGAs and utilizes less than 9% of the BRAMS in a Xilinx Spartan6 LX75 FPGA, used in current production vehicles for stereo vision. The information, which average disparity and which standard deviation occurs at every pixel is stored in the discretized versions of the average disparity image and standard deviation image externally. These images are read in every frame from RAM (10bits · 112640pixels) which adds just 30Mbit/s additional bandwidth to the system at 25Hz. The average disparity value and the standard deviation is read in for every pixel and these values are used for a lookup to the corresponding cost vector that is added to all 64 disparity hypotheses at the respective pixel. An overview of the new cost and prior computation block is shown in 6.

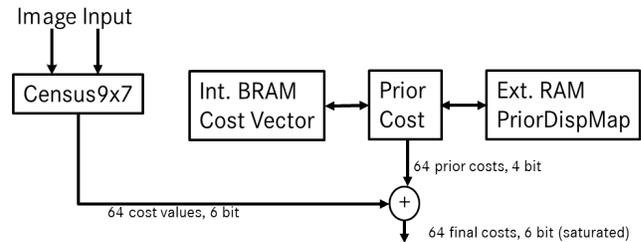


Figure 6. Overview of the data cost computation block using the scene prior on FPGA.

The total resource consumption is 15 BRAMs for the scene prior cost vector storage, about 100 lookup tables (LUTs) for the cost addition and the logic for reading in from RAM, and about 30Mbits/s additional bandwidth. This totals to a very mild increase in resources in all categories resulting in a significant performance increase as shown next. No additional latency is introduced.

## 6. Results

### 6.1. Results on the KITTI Dataset

We tested our approach on the KITTI data set. However, no adverse weather conditions are present in the data, so we can only verify that the scene prior does not decrease the performance. The only parameter to be adjusted is the outlier rate  $p_{out}$ .

$p_{out}$	SGM $P_{scene}$ error [%]	SGM $P_{scene}$ density [%]
base	6.01	94.5
0.95	6.00	95.1
0.90	6.04	95.1
0.85	6.22	95.0
0.80	6.22	95.0
0.50	6.77	94.7
0.20	9.01	93.3

Table 1. Erroneous pixel rates on the KITTI training set. Erroneous pixels deviate by more than 3px to the ground truth (KITTI default), stereo density refers to the percentage of valid disparities.

We evaluated on the full training set. The stereo data is interpolated according to the KITTI development kit. Table 1 summarizes the results. We observe little difference in false pixel statistics. For the disparity density we see little difference as well with slight advantages to using a scene prior. Outlier rates below 80% result in noticeable higher erroneous pixel rates.

## 6.2. Results on the Ground Truth Stixel Dataset

An appropriate database for testing our scene prior must contain adverse weather scenarios. The Ground Truth Stixel Dataset [4] fulfills this requirement. It contains mostly rainy highway scenes with blurred windshield, wiper passes, and spray water behind passing cars. A few good weather scenes are included to make sure no undesired effects on normal scenes occur.

A percentage of false positive points is computed using the driving corridor, an area in the scene directly in front of the car,  $2m$  in width and  $2m$  in height. The corridor is calculated from the car’s current position using ego-motion data, describing an area that the car covers one second into the future. This volume is free of any protruding structures; any triangulated world point that lies within the corridor is counted as a false positive point.

In addition, the data is annotated with an intermediate representation called stixels, thin rectangular sticks with upright orientation that contain image positions and disparities for objects or structures. This way, about 10% of the disparities in the image are labeled. The stixel algorithm as introduced in [23] is computed and our stereo algorithm variants serve as input. A stixel example image with input stereo image (left) is shown in Figure 7 center. This stixel representation encodes the necessary information for most driver assistance functions. Stixels that occupy the driving corridor volume are false positives. To compute a detection rate for the stixels, we use ground truth data that covers a distance of about  $50m$ , counting all stixels within  $3\sigma_d$  to the ground truth as true positives.

Table 2 shows the results on this 3000-frames database.

	false positive		detection
	point rate [%]	stixels [#frames]	rate [%]
SGM	0.23	261	85.7
GC	0.25	316	85.7
iSGM	0.22	182	84.4
SGM SimplePrior [13]	0.04	80	83.5
SGM conf [4]	n.a.	45	80.2
GC SimplePrior [13]	0.05	123	83.7
SGM ScenePrior(0.95)	0.20	236	85.3
SGM ScenePrior(0.9)	0.09	136	84.8
SGM ScenePrior(0.85)	0.06	108	83.8
SGM ScenePrior(0.8)	0.05	91	82.7
SGM ScenePrior(0.5)	0.02	57	77.0
SGM ScenePrior(0.2)	<b>0.01</b>	<b>40</b>	70.7
GC ScenePrior(0.95)	0.17	263	<b>86.1</b>
GC ScenePrior(0.9)	0.11	197	84.0
GC ScenePrior(0.85)	0.09	152	82.6
GC ScenePrior(0.8)	0.07	113	81.0
GC ScenePrior(0.5)	0.03	81	73.9
GC ScenePrior(0.2)	0.01	65	68.4

Table 2. Comparison of false positive point rates, number of frames with false positive stixels, and detection rates on the Ground Truth Stixel Database.  $p_{out}$  in parentheses for the scene prior.

All considered algorithms use the same Census data term. As a baseline, SGM and GC are shown in the first two lines. GC performs slightly worse, probably due to the Median filter in SGM post-processing. As additional baseline serves iterative SGM introduced in [24] which performed best on the Robust Vision Challenge<sup>1</sup>. It delivers slightly better results than basic SGM. With the simple scene prior (SimplePrior) from [13], false positive point rate and false positive stixel numbers drop by more than a factor of two. With our pixel-based scene prior (ScenePrior), we also see a clear reduction of false positives. The results for different outlier probabilities are shown with a sweet spot around 0.8 where false positive rates drop by a factor between 3 and 5 and detection rate is only mildly degraded. Note the comparable performance of SGM and GC with the scene prior. From this we can see that it is best to adapt the outlier rate to the current weather conditions: The worse the weather conditions, the lower the outlier rate in order to maintain a low false positive level. With our scene prior below  $p_{out} = 0.8$  we perform similar on false positive level to [4], where stereo confidence information is used in addition (SGM conf).

Note that for the simple scene prior used in [13], the false positive rate is reduced at the expense of losing stereo infor-

<sup>1</sup><http://hci.iwr.uni-heidelberg.de/Static/challenge2012/>



Figure 7. Input disparity image (left) and resulting stixel representation (middle). The labeled ground truth is shown in blue on the right, the corridor in red.

	false positive		detection rate [%]
	point rate [%]	stixels [#frames]	
rain and night	0.53 (3.15)	105(308)	91.3 (92.2)
rain and day	0.27 (0.76)	24 (60)	61.2 (67.1)
snow and day	0.03 (0.12)	1 (10)	95.9 (94.1)

Table 3. false positive point rates, number of frames with false positives, and detection rates for different weather scenarios using SGM ScenePrior (SGM baseline in parentheses) .

mation on the street necessary for street surface estimation (see Figure 8) which does not happen with the proposed scheme.

### 6.3. Results for a Night/Rain/Snow Dataset

We also applied our scene prior to more types of adverse weather conditions (night, rain, snow) with 12bit/pixel imagery computing 128 disparity steps on 1200 frames. Figure 8 top shows results for a night and rain scene just before the windshield wiper passes. The basic SGM is shown on the left, SGM with the simple scene prior [13] in the middle and the new scene prior on the right. The red blobs above the car disappear and the holes in the street are mostly filled, in contrast to the simple scene prior where all street information is lost. In the bottom row a snow result is shown.  $p_{out}$  was set to 0.8. The rain scene result in Figure 1 is computed with the same setting.

We annotated some parts in above challenging scenes (rain, snow, night and rain) with ground truth (mainly cars and traffic signs) and used the ground truth stixel dataset evaluation scheme. Table 3 summarizes the results for different scenarios comparing SGM with scene prior to the SGM baseline. The false positive rate drops dramatically while the detection rate remains constant.  $p_{out} = 0.8$  is used.

For reference we show the averaged results over these scenes for different outlier rates in Table 4. Here,  $p_{out}$  around 0.8 also appears to be the best compromise between detection rate and false positive rate. The detection rate of the simple scene prior is low in comparison to the scene prior variants shown here.

	false positive		detection rate [%]
	point rate [%]	stixels [#frames]	
SGM	1.34	378	83.5
SGM SimplePrior [13]	<b>0.05</b>	24	77.2
SGM ScenePrior(0.95)	0.82	323	<b>84.6</b>
SGM ScenePrior(0.9)	0.44	212	82.9
SGM ScenePrior(0.85)	0.31	145	81.9
SGM ScenePrior(0.8)	0.25	124	81.2
SGM ScenePrior(0.5)	0.10	39	77.5
SGM ScenePrior(0.2)	0.07	<b>18</b>	75.9

Table 4. Comparison of false positive point rates, number of frames with false positives, and detection rates on Night/Rain/Snow Scenes (integrated over 3 scenes).  $p_{out}$  in parentheses for the scene prior.

## 7. Conclusions and Future Work

We have presented a scene prior, incorporated both into Graph Cut and Semi-Global Matching that is able to reduce false positive rates in driver assistance scenarios while maintaining detection rate. Along the way, we obtained disparity statistics for typical traffic scenes. The probabilistic problem formulation allowed us to integrate the prior efficiently into the data term applicable to any stereo algorithm generating a matching cost volume. We showed efficient ways to implement the scene prior both on CPU and FPGA systems.

For future work, combining the prior with stereo confidence is a promising line of research. Also, we will explore the option to adapt the outlier rate of the scene prior to the weather conditions.

## References

- [1] D. Scharstein and R. Szeliski, "Middlebury online stereo evaluation," viewed 2013/03/12, <http://vision.middlebury.edu/stereo>. 1
- [2] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Int. Conference on Computer Vision and Pattern Recognition 2012*, June 2012. 1
- [3] S. Meister, B. Jaehne, and D. Kondermann, "Outdoor stereo camera system for the generation of real-world benchmark

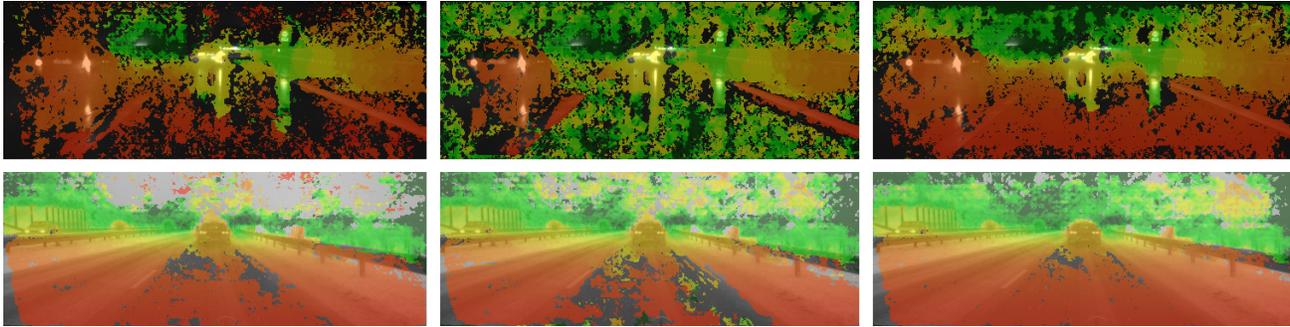


Figure 8. Night-time and rain traffic scene (top) and snow scene (bottom). Stereo reconstruction (red=near ... green=far) for the scene using SGM (left), SGM with simple scene prior (center) and with the new scene prior (right). Large red blobs indicate nearby objects leading to potential false positive objects.

- data,” *Journal of Optical Engineering*, vol. 51, no. 2, 2012. 1
- [4] D. Pfeiffer, S. Gehrig, and N. Schneider, “Exploiting the power of stereo confidences,” in *Int. Conference on Computer Vision and Pattern Recognition 13*, June 2013. 1, 6
- [5] H. Hirschmuller, “Accurate and efficient stereo processing by semi-global matching and mutual information,” in *Int. Conference on Computer Vision and Pattern Recognition 05, San Diego, CA*, vol. 2, June 2005, pp. 807–814. 1, 2
- [6] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” in *Proceedings of Int. Conference on Computer Vision 99*, 1999, pp. 377–384. 2
- [7] R. Ranftl, S. Gehrig, T. Pock, and H. Bischof, “Pushing the limits of stereo using variational stereo estimation,” in *Intelligent Vehicles Symposium 2012*, June 2012. 2
- [8] S. Hermann and R. Klette, “Inclusion of a second-order prior into semi-global matching,” in *PSIVT*, 2009. 2
- [9] S. Gehrig, H. Badino, and U. Franke, “Improving sub-pixel accuracy for long range stereo,” *Computer Vision and Image Understanding (CVIU)*, vol. 116, no. 1, pp. 16–24, January 2012. 2
- [10] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *Int. Conference on Computer Vision and Pattern Recognition 99*, 2006, pp. 519–528. 2
- [11] K. Kofuji, Y. Watanabe, T. Komuro, and M. Ishikawa, “Stereo 3d reconstruction using prior knowledge of indoor scenes,” in *Proceedings of the IEEE Conference on Robotics and Automation 11*, 2011, pp. 5198–5203. 2
- [12] D. Gallup, J. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, “Real-time plane-sweeping stereo with multiple sweeping directions,” in *Int. Conference on Computer Vision and Pattern Recognition 07*, 2007. 2
- [13] S. Gehrig, M. Reznitskii, N. Schneider, U. Franke, and J. Weickert, “Priors for stereo vision under adverse weather conditions,” in *Computer Vision for Autonomous Driving@ICCV*, 2013. 2, 6, 7
- [14] H. Hirschmuller and D. Scharstein, “Evaluation of stereo matching costs on images with radiometric distortions,” *IEEE Transact. Pattern Analysis & Machine Intelligence*, vol. 31, no. 9, pp. 1582–1599, 2009. 2
- [15] R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” in *ECCV*, 1994, pp. 151–158. 2
- [16] H. Hirschmuller and S. Gehrig, “Stereo matching in the presence of sub-pixel calibration errors,” in *Int. Conference on Computer Vision and Pattern Recognition 09, Miami, FL*, June 2009. 2
- [17] I. Haller and S. Nedeveschi, “GPU optimization of the SGM stereo algorithm,” in *ICCP*, 2010. 2
- [18] A. Delong, A. Osokin, H. Isack, and Y. Boykov, “Fast approximate energy minimization with label costs,” *IJCV*, vol. 96, no. 1, pp. 1–27, 2012. 2, 5
- [19] N. Soquet, D. Aubert, and N. Hautiere, “Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation,” in *IV 2007*, 2007. 3
- [20] R. Gray, *Entropy and information theory*. Springer-Publishing New York, 1990. 5
- [21] S. Gehrig, F. Eberli, and T. Meyer, “A real-time low-power stereo vision engine using semi-global matching,” in *ICVS 2009*, October 2009, pp. 134–143. 5
- [22] J. I. Woodfill and et al., “The tyzx deepsea g2 vision system, a taskable, embedded stereo camera,” in *Embedded Computer Vision Workshop*, 2006, pp. 126–132. 5
- [23] D. Pfeiffer and U. Franke, “Towards a global optimal multi-layer stixel representation of dense 3d data,” in *BMVC*, September 2011. 6
- [24] S. Hermann and R. Klette, “Iterative SGM for robust driver assistance systems,” in *ACCV*, 2012. 6