Priors for Stereo Vision under Adverse Weather Conditions

Stefan Gehrig[1] Maxim Reznitskii[2] Nicolai Schneider[3] Uwe Franke[1] Joachim Weickert[2]
[1] Daimler AG, HPC 050 G 024, 71059 Sindelfingen, Germany
[2] Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
[3] IT-Designers, Entennest 2 73730 Esslingen, Germany

Abstract

Autonomous Driving benefits strongly from a 3D reconstruction of the environment in real-time, often obtained via stereo vision. Semi-Global Matching (SGM) is a popular method of choice for solving this task which is already in use for production vehicles. Despite the enormous progress in the field and the high level of performance of modern methods, one key challenge remains: stereo vision in automotive scenarios during weather conditions such as rain, snow and night. Current methods generate strong temporal noise, many disparity outliers and false positives on a segmentation level. They are addressed in this work. We formulate a temporal prior and a scene prior, which we apply to SGM and Graph Cut. Using these priors, the object detection rate improves significantly on a driver assistance database of 3000 frames including bad weather while reducing the false positive rate. We also outperform the ECCV Robust Vision Challenge winner, iSGM, on this database.

1. Introduction

Stereo 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 [26]. The Middlebury database [26] is a good resource of available stereo algorithms, but its scene complexity is limited. A more challenging benchmark is the KITTI database [9], comprising of some 200 image pairs of street scenes. It still under-represents 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 scenarios are the focus of our work.

In the light of increasing autonomy of future vehicles, such scenarios have to be mastered. Work on benchmarking such scenarios has just recently started. The Heidelberg HCI dataset [1] was the first data set covering challenging weather scenarios, however, without supplying ground truth. The Ground Truth Stixel Dataset [23] contains a set of rainy highway scenes with sparse ground truth labels for the free space.

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. A successful choice for solving this challenging task in real-time is Semi-Global Matching [14] (SGM), which can also be found in the top 10 of the KITTI benchmark.

Even SGM cannot measure image parts that are, for example, occluded by the windshield wiper, although the scene was clearly measured in the previous frame. Also, SGM has a uniform disparity distribution for outlier disparities which automatically generates a peak towards nearby 3D points.

To counteract against these two observations we introduce two types of priors:

Temporal prior: In image sequence analysis most of the scene content is seen several times when operating at high frame rates. We use the 3D reconstruction obtained via camera geometry and disparity map from the previous frame, predict it into the future considering the ego-motion and assuming a predominantly static world, and use the result as a temporal prior.

Scene prior: When looking at outlier disparity distributions, *e.g.* in the KITTI database, we see a roughly uniform distribution. Due to the hyperbolic nature of the disparity-distance relation, the disparity outliers occur more frequently in the nearby 3D volume. Unfortunately, this is the area where false positives hurt the most: right in front of the ego-vehicle where immediate action has to be taken to avoid collisions. We counterbalance this effect by introducing a 3D scene prior that expects a more uniform distribution in 3D space.

In this paper, we first transfer the excellent engineering approach of SGM into a probabilistic formulation, maintaining successful similarity measures and smoothness parameters. Then, we integrate new priors to improve stereo, solve the optimization problem with Graph Cut [2] (GC),

[1] http://hci.iwr.uni-heidelberg.de/Benchmarks

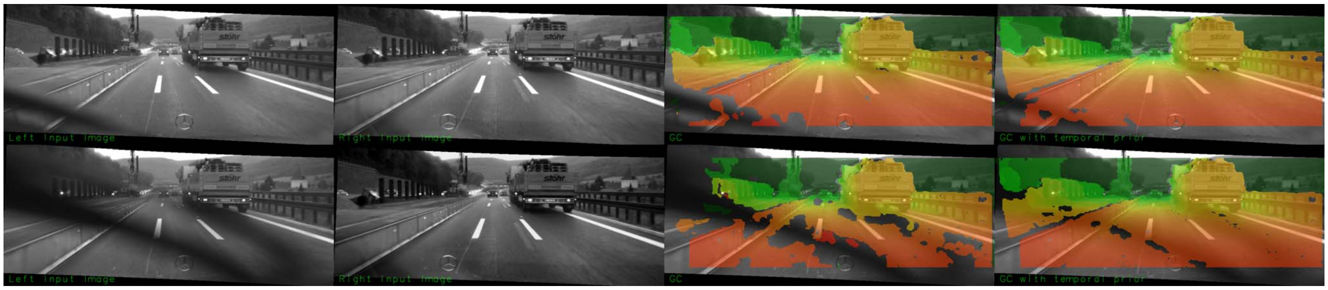

Figure 1. Two consecutive frames (top and bottom) of a rainy highway scene with windshield wiper passing. From left to right, the images are: left image, right image, color-coded stereo reconstruction (red=near ... green=far, no color=invalid) using GraphCut, and stereo using Graph Cut with priors. Large red blobs indicate nearby objects that lead to potential false positive objects. Figures are best viewed in color.

and transfer the priors back to SGM. Transferring SGM into a probabilistic formulation helps in understanding how to incorporate the priors in a Bayesian sense. Summarizing, the main contributions of this paper are: the transfer of SGM into a Bayesian Framework and solving it via GC; the introduction of efficient and effective temporal and scene priors, 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. We also perform the transfer of the new priors back into SGM, maintaining efficiency.

The rest of the paper is organized as follows. Section 2 covers related work in stereo to incorporate priors. Section 3 covers our GC approach driven by a successful SGM engine, embedding it in a probabilistic framework. In Section 4 we detail how to incorporate the priors in this Bayesian framework and how to implement them efficiently. Section 5 shows results for GC, SGM and the new priors 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 stereo variants introduced in this paper.

2. Related Work

We limit ourselves to related work in stereo vision using priors or temporal stereo. "Prior" in the context used here means prior information that is independent of the image information in the current stereo pair. We do not address the options of smoothness priors. In [6], 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. Scene priors using merely the distribution of disparity values have not been used before. Shape priors are popular in multi-view stereo (see *e.g.* [27]). Planarity and orthogonality have been exploited as priors several times, *e.g.* in [18].

Temporal priors have been investigated previously. Most work covered so-called scene flow approaches, where disparity and flow are estimated simultaneously based on the current and previous image pair, *e.g.* [17]. Similarly, space-time stereo uses multiple time steps to simultaneously estimate position and motion [3]. A fast version is 6D-Vision, introduced in [5] and which has also been applied to dense data [24]. In contrast to these approaches, we do not attempt to estimate both stereo and motion but want to use 3D information from previous time steps to support the current stereo computation. There are also several works in optical flow estimation using priors, such as the assumption of a static world [20], [29]. Classic works in optical flow already exploited more than two frames several decades ago [1], [21].

The problem of temporally consistent stereo has recently been addressed several times. Larsen *et al.* [19] use belief propagation in space and time to generate a temporal consistent stereo result. The method is designed for view interpolation in offline use and runs 15 minutes per frame. Reichardt *et al.* [25] use an adaptive support weight stereo approach with bilateral filtering, also in the time domain. The method runs fast on a GPU with high memory demands.

The method closest to our temporal prior approach is the work of Gehrig *et al.* [7]. It uses an iconic Kalman filter to filter the disparities, assuming a static world and predicting the data with known ego-motion. A disparity map is generated in one step and modified in a second step. Our method also assumes a static world and makes use of ego-motion, but the temporal prior is applied as part of a single disparity optimization step and hence the algorithm is able to recover from ambiguous matches resulting in better performance under adverse conditions (see Line 5 in Table 2).

3. Semi-Global Matching vs. Graph Cut

3.1. Semi-Global Matching

Roughly speaking, SGM [14] 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).

Many stereo algorithms favor the choice of an adaptive smoothness penalty that depends on the image gradient (*e.g.* [14]) because it gives better object boundaries. We keep our smoothness penalty constant because it performs better on bad weather data (see Line 6 in Table 2).

Hirschmueller *et al.* [16] achieved very good performance SGM using the Hamming distance of images transformed with a 9x7 Census as similarity criterion [30]. Other investigations (*e.g.* [15], [11]) have confirmed this finding. We apply this similarity criterion throughout our work.

In order to identify and exclude occlusions, Hirschmueller [14] 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.* [2]. For our investigations, we use the implementation from [4] 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.

3.3. Drawing Parallels Between Semi-Global Matching with Graph Cut

We apply the same smoothness potential to GC as is used in Semi-Global Matching in order to model slanted surfaces such as the street well. However, good parameter choices for SGM ($P_1 = 20, P_2 = 100$) represent only a semi-metric since they violate the condition: $P_2 \leq 2P_1$. We use the alpha-expansion algorithm and approximate the original potential function, assigning intermediate penalties for deviations between 2 and 4 pixel in our case. As an alternative, one can apply the swap-label algorithm since it is able to deal with semi-metrics. We compared the two options, obtained slightly better performance and faster runtime with the alpha-expansion version, and use this variant for the remainder of this paper.

We also use the SGM data cost function. The Hamming distance HAM of the Census-transformed images can be represented in a Bayesian approach as a probability distribution:

$$p(c_{li}, c_{ri}|d_i) \propto e^{-\lambda HAM(c_{li}, c_{ri-d_i})}, \tag{1}$$

where c_{li}/c_{ri} represents the Census string for the i-th pixel in the left/right image and d_i the disparity at that pixel. To achieve comparable performance without changing the SGM parameters, one has to multiply the GC data costs by the number of SGM paths (here, 8). Also, the connectivity of the graph in GC has to be increased to an 8-neighborhood to reflect the diagonal paths of SGM. Note that we do **not** attempt to model the independent 1D paths of SGM in a probabilistic framework. Instead, we maintain the same data and smoothness terms in GC and assume that SGM approximates the 2D optimum well.

Examining image pairs that have visually no outliers in the disparity maps, we actually obtain a distribution of Hamming distances for the correct disparities that closely resembles a Laplacian distribution with $\lambda = 4.58$ and 1.3% standard error of the fit. Similar results are obtained from the KITTI training set [9] with known ground truth. This gives us the weighting term between the Census costs and the upcoming stereo priors and confirms the assumption of the probability distribution of the data costs being Laplacian.

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 with a minimal visual advantage for GC. This can be expected since GC performs a full 2D optimization. An example is shown in Figure 2.

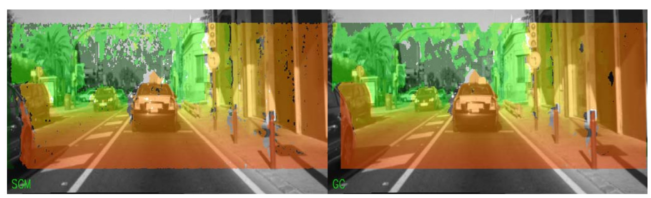

Figure 2. Standard traffic scene overlaid with disparity result SGM (left) and GC right. Red pixels are near, green far away.

4. Temporal Prior and 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. Additional priors are able to stabilize disparity maps in such situations.

4.1. Incorporation of the Priors

To understand how the priors are incorporated we describe our stereo task in a probabilistic fashion and extend it with the new priors. We seek the disparity map D that maximizes the probability

$$p(D|\hat{D}, I_L, I_R, P_{scene}) \propto p(\hat{D}, I_L, I_R, P_{scene}|D) \cdot p(D), \tag{2}$$

where I_L/I_R is the left/right image, $\hat{D}$ the predicted disparity map from the previous frame, and P_{scene} the scene prior. $p(D)$ represents the smoothness term.

Decomposing this with the assumption of the image data and the priors to be independent we obtain

$$p(\hat{D}, I_L, I_R, P_{scene}|D) \propto p(I_L, I_R|D) \cdot p(\hat{D}|D) \cdot p(P_{scene}|D). \tag{3}$$

The first term is the Census data term, the second one the temporal prior term and the third one the scene prior term. The first term has been introduced in the previous section, the latter two are detailed in the following. All above terms are carefully normalized to one to obtain their relative weights automatically, without additional parameter tuning.

4.2. Temporal Prior

The temporal prior addresses situations where the disturbances of the optical path change frequently due to image noise (night), atmospheric effects (rain, snow) or blockage (wiper passes). More persistent effects such as fog, backlight and reflections are not addressed.

The assumption of a known ego-motion makes the temporal prior very efficient to implement since every disparity is only dependent on the predicted disparity at the same image location:

$$p(\hat{D}|D) = \prod_i p(\hat{d}_i|d_i) = \prod_i ((1-p_{out})\mathcal{N}(d_i, \sigma_d) + p_{out} \cdot \mathcal{U}). \tag{4}$$

$\mathcal{N}$ is the normal distribution of $\hat{d}_i$ with mean d_i and standard deviation σ_d as parameters ($\mathcal{N} \propto e^{-(d_i - \hat{d}_i)^2/2\sigma_d^2}$), $\mathcal{U}$ the uniform distribution, and p_{out} is the outlier probability for the prediction to be wrong, due to wrong disparity, wrong ego-motion, or violation of the static-world assumption. This mixture model of Gaussian distribution and uniform distribution has been observed frequently when comparing to ground truth data (*e.g.* [22]).

To obtain our predicted disparity, we use the camera parameters to project every disparity from the previous frame into the world and transform the 3D world point into the current camera coordinate system using ego-motion information. Then, the point is projected back into the image with a new disparity value entered in the nearest image coordinate. This produces a predicted disparity map with tolerable holes due to the warping step (Figure 3 top left). In adverse weather conditions, a high percentage of outliers occur. Most of them can be removed by applying the RL-consistency check. This check is applied *b*efore the warping step to reduce the number of erroneous disparities to be warped.

4.2.1 Using Flow Information to Obtain Static-World Confidence

Applying this temporal prior term to our algorithm as is would cause deterioration in scenes with moving objects. Similar to [20], which uses ego-motion and a static-world assumption to initialize optical flow, we use sparse optical flow measurements to find regions where this static-world assumption fails. However, we do not follow the idea of using flow directly since the results are too unstable in bad weather scenes. Instead, a check of the static-world assumption via optical flow is performed to determine a temporal prior confidence c_{tp} (Figure 3 bottom right).

With the assumption of a static world, we can compute an optical flow field using just ego-motion and disparity information (Figure 3 top right). We can also compute optical flow in the classical sense, using the previous and the current grayvalue images as inputs (Figure 3 top center). If the world is static, the two methods should produce compatible results.

A comparison of the outputs from the above two methods yields the flow residual image R_f (Figure 3 bottom left). This information is used for the temporal prior extending $p(\hat{D}|D)$ to $p(\hat{D}, R_f|D)$ maintaining above decomposition properties. The residual flow is used to determine a confidence c_{tp} for the temporal prior, replacing σ_d in Equation 4 with σ_d/c_{tp}, confirmed by statistics from traffic scene data. We explicitly try to keep the influence of optical flow information small since we want to address difficult situations where flow is even harder to measure than stereo. Note that, for the RL-consistency check, we need to compute optical flow two times: once on the left and once on the right image.

In our implementation, we use a sparse, robust and efficient optical flow method [28] and apply a 5x5 dilation to densify it (Figure 3 top middle). If areas have no flow measurement, we set $c_{tp} = 0.9$ to allow more smoothing. Before generating the confidence image we enter inconsistent flow vectors with more than 2px deviation into the confidence map with $c_{tp} = 0$, reflecting the fact that these regions are likely to be flow occlusions and hence cannot be measured. The other flow residuals (r_f) are converted to confidences c_{ts,r_f} via

$$c_{tp,r_f}(i) = e^{-r_f^2(i)/\sigma_{r_f}^2}. \tag{5}$$

We choose $\sigma_{r_f} = 2$ to reflect flow inaccuracy. An example confidence image is shown in Figure 3 bottom right (white represents high confidence).

4.2.2 Considering Ego-Motion Uncertainties

So far we assumed perfect ego-motion. Slight rotational ego-motion errors are noticeable especially at depth edges. In order to efficiently deal with such ego-motion errors, we

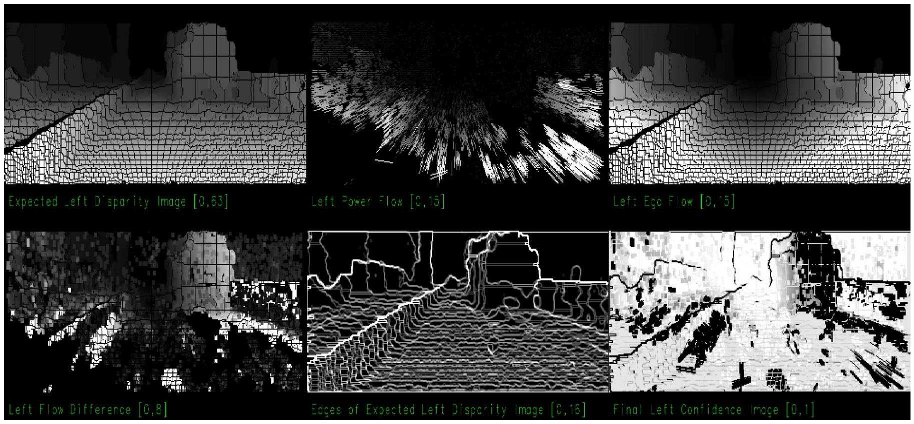

Figure 3. Highway scene from Figure 1 with intermediate results for temporal prior: warped disparity image due to ego-motion (top left), optical flow result (top middle), computed flow with static-world assumption (top right), residual flow based on above flows (bottom left), disparity gradient of the predicted disparity (bottom middle) and final temporal prior confidence image (bottom right). The resulting disparity image is shown in Figure 1 on the right. Note the fewer outliers compared to classic GC stereo in Figure 1.

also reduce the temporal prior confidence at depth discontinuities:

$$c_{tp,dd}(i) = \frac{1}{\pi} \cdot (arctan(d_{slope} \cdot (\nabla d_{ref} - |\nabla d_i|)) + \frac{\pi}{2}), \quad (6)$$

where ∇d_{ref} is a constant representing the acceptable disparity gradient for smoothing (1 in our case), $d_{slope} = 5$ determines the slope of the confidence, and $c_{tp,dd}$ is the confidence based on disparity edges. This parametrization ensures that slanted surfaces such as the street and fronto-parallel surfaces are still temporally smoothed but the regions around disparity discontinuities are not smoothed. A 3x3-Sobel operator is used for the gradients which leads to $3px$ margins that are not smoothed at depth discontinuities. A depth edge detection result is shown in Figure 3 bottom middle. The final confidence is just a multiplication of the single parts: $c_{tp} = c_{tp,r_f} \cdot c_{tp,dd}$ (Figure 3 bottom right).

4.3. Scene Prior

Many types of scene priors are conceivable and some have been discussed in the related work section. We choose a simple prior that assumes that all observed world points are equally distributed in 3D space ($p(z) = const.$). The perspective projection counterbalances this effect towards a uniform disparity distribution, so we use disparity data from highway scenes to fit a disparity distribution that can model both uniform disparity distributions ($\alpha = 0$ in Equation 7) and uniform distance distribution ($\alpha = 1$).

Like the temporal prior, the scene prior can also be decomposed pixel-wise. For every pixel, the scene prior probability is computed with

$$p(P_{scene}|d_i) \propto \frac{1}{((d_i + 1)d_i)^{\alpha}} \; for \; d_i \geq d_{min}. \quad (7)$$

Due to normalization problems, one has to limit this assumption to a finite range, a maximum distance. For distances corresponding to disparities smaller than d_{min}, the prior probability is set to $p(P_{scene}|d_{min})$. α is determined to be 0.29 with 10% standard error of the fit. Again, this only adds a unary term to the data cost volume and is efficient to implement. Roughly speaking, we introduce a slight slant towards small disparities into the data term to prefer small disparity values in ambiguous cases. It is straightforward to add more complex and pixel/region-specific priors (*e.g.* sky on top of the image, street and/or objects in the lower part) in the same way.

All probabilities introduced above are transferred to log-likelihood energies [10] and these energies are fed into the GC engine from [4]. The priors are easily transferred back to SGM since only the data term is affected.

5. Results

5.1. Results on the KITTI Dataset

We tested our approach on the KITTI dataset. However, no adverse weather conditions are present in the data, so we can only verify that the temporal and scene prior does not decrease the performance. In addition, the basic stereo benchmark does not include stereo sequences but 10 additional frames before the evaluated stereo pair are available in the multi-view extension without ego-motion data. We used this data and determined ego-motion crudely assuming only longitudinal motion for the selected scenes.

We limit our evaluation to interesting examples of the training set focusing on moving objects (cars, pedestrians, cyclists) that violate our static world assumption. One simple scene (image pair 162) is included for reference. The stereo data is interpolated according to the KITTI development kit. Table 1 summarizes the results. We see little difference in false pixel statistics with advantages to using a temporal prior, similar for the disparity density before back-

image pair number	162	045	057	164
SGM (error)	6.2%	8.6%	18.3%	7.0%
SGM temp prior	5.8%	8.2%	14.8%	6.2%
SGM (density)	93.7%	90.4%	92.3%	91.9%
SGM temp prior	95.7%	92.6%	94.9%	93.8%

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

ground interpolation.

5.2. Results on the Ground Truth Stixel Dataset

An appropriate database for testing our priors must contain adverse weather scenarios. The Ground Truth Stixel Dataset [23] advertised on the KITTI web page 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.

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 [22] is computed and our stereo algorithm variants serve as input. A stixel example image with input stereo image (left) is shown in Figure 4 center.

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; those stixels that occupy the volume are false positives. Similarly, any triangulated world point that lies within the corridor is counted as a false positive point. 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. All considered algorithms use the same Census data term as explained in Section 3.3. 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 [13] which performed best on the Robust Vision Challenge[2]. It delivers slightly better results than basic SGM. Baselines that exploit temporal information such as 6D-Vision[24] and stereo integration [7], fed with the SGM baseline, can outperform the SGM baseline but fall back clearly to the priors introduced here. With the scene prior (ScenePrior), both false positive point rate and false positive stixel numbers drop by factor of five. With a temporal prior (TempPrior), we see both a reduction of false positives and an increase in detection rate. Much of the increase in detection rate is attributed to the fact that this prior helps inpainting old stereo data when the windshield wiper passes (rightmost column). Additionally applying the scene prior (TempScenePrior) leads to even lower false positive rates while maintaining excellent detection rate. Note the similar performance of SGM and GC with priors. This combination of both priors yields the lowest false positive stixel numbers and performs similar to [23], where stereo confidence information is used in addition (SGM conf) at a much higher detection rate. This confidence is obtained from the quality of the disparity sub-pixel fit.

For the temporal prior, we set $\sigma_d = 1$ and $p_{out} = 0.2$, a slightly pessimistic outlier rate determined from false positive point statistics in very challenging scenarios. We set $d_{min} = 5$ for the scene prior (*i.e.* $50m$ maximum distance). Note that all parameters choices correspond to sensor properties (disparity noise, optical flow noise, measurement range) or are determined from statistics.

5.3. Results for an Adverse Weather Dataset

We also applied our new priors to several types of adverse weather conditions with 12bit/pixel imagery at 25Hz computing 128 disparity steps. Figure 5 shows results for a night and rain scene just before the windshield wiper passes. The basic GC is shown on the left, GC with temporal smoothness in the middle and the combination of temporal and scene prior on the right. The red blobs above the car disappear and the holes in the street are partially filled with priors. In Figure 6 we show a result for SGM and for SGM with priors.

We annotated some parts in above challenging scenes with ground truth (mainly cars and traffic signs) and determined a false positive point rate and detection rate as described above. Table 3 summarizes the results for different scenarios comparing SGM with both priors to the SGM baseline. The false positive rate drops dramatically while the detection rate is even higher.

The computational overhead to compute the priors is small. With little optimization and the maximum approximation for the disparity distribution (*e.g.* [12]), the runtime is $30ms$ additional computation time for both stereo priors, excluding optical flow which is computed on an FPGA. Timings are for a Core(TM) i7 PC on images downscaled by a factor of two, resulting in 512x160px.

[2]http://hci.iwr.uni-heidelberg.de//Static/challenge2012/

Figure 4. 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 points/stixels			detection rates	
	% fp point	#fp	#fp frames	all %	w/ wiper %
SGM	0.23	1243	261	85.7	60.2
GC	0.25	2211	316	85.7	61.2
iSGM	0.22	3072	182	84.4	65.0
6D-Vision	0.20	923	166	84.0	51.8
StereoIntegration	0.38	5442	945	82.4	82.4
SGM AdaptiveP2	0.50	1439	546	82.9	53.0
SGM conf [23]	n.a.	107	45	80.2	n.a.
GC ScenePrior	0.05	375	123	83.7	57.2
GC TempPrior	0.04	652	165	88.4	**82.6**
GC TempScenePrior	0.01	**143**	**59**	87.3	80.2
SGM ScenePrior	0.04	220	80	83.5	56.2
SGM TempPrior	0.03	658	141	**88.5**	82.6
SGM TempScenePrior	**0.01**	229	63	87.0	78.6

Table 2. Comparison of false positive point rates, false positive stixels (fp), number of frames with false positives (fp frames), and detection rates on the Ground Truth Stixel Database. The combination of both priors yields the best overall result for SGM and GC.

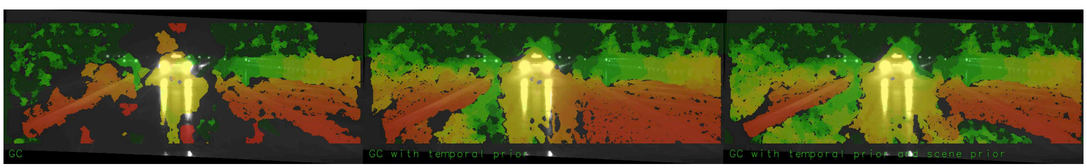

Figure 5. Night-time and rain traffic scene. Stereo reconstruction (red=near ... green=far) for the scene using GC (left), GC with Temporal Prior (center) and with both priors (right). Large red blobs indicate nearby objects leading to potential false positive objects.

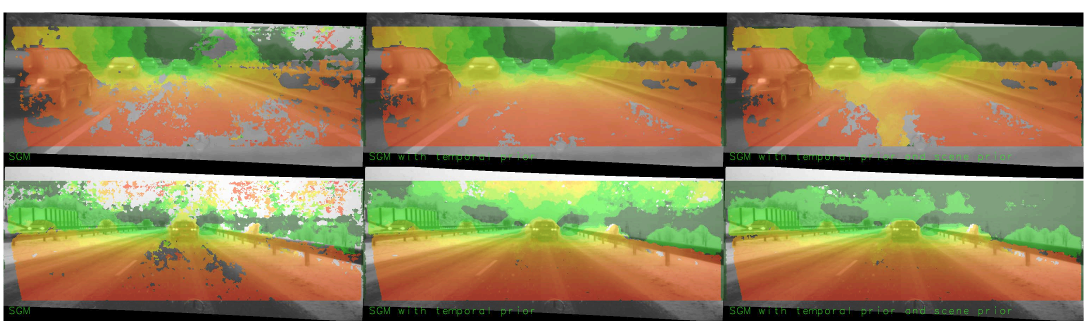

Figure 6. Rain scene (top) and snow scene. Stereo reconstruction (red=near ... green=far) for this scene using SGM (left), SGM with Temporal Prior (center) and with both priors (right). Large red blobs indicate nearby objects leading to potential false positive objects.

6. Conclusions and Future Work

We have presented a temporal and scene prior, incorporated both into Graph Cut and Semi-Global Matching that is able to reduce false positive rates in driver assistance scenarios while even increasing detection rate. The probabilistic problem formulation allowed us to integrate the priors efficiently into the data term applicable to any stereo algorithm.

Future work will further explore the idea of scene priors, e.g. on per-pixel level. In addition, we want to integrate the shown priors in real-time systems (*e.g.* [8]) to enhance autonomous driving. Finally, combining the priors with stereo confidence is another promising line of research.

	false positive points/stixels			detection rates	
	% fp point	#fp	#fp frames	all %	w/ wiper %
rain and night	0.02 (3.15)	24 (2502)	5 (308)	95.5 (92.2)	n.a.(n.a.)
rain and day	0.04 (0.76)	26 (279)	9 (60)	76.2 (67.1)	65.3 (38.5)
snow and day	0.00 (0.12)	1 (18)	1 (10)	95.2 (94.1)	97.9 (64.4)

Table 3. false positive point/stixel rates (fp) and detection rates for different scenarios using SGM TempScenePrior - SGM baseline in () .

References

[1] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Int. Conference on Computer Vision and Pattern Recognition 91*, pages 292–302, 1991. 2

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *Proceedings of Int. Conference on Computer Vision 99*, pages 377–384, 1999. 1, 3

[3] J. D. Davis, D. Ramamorrthi, and S. Rusinkiewicz. Spacetime stereo: a unifying framework for depth from triangulation. *PAMI*, 27:296 – 302, 2005. 2

[4] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *IJCV*, 96(1):1–27, 2012. 3, 5

[5] U. Franke, C. Rabe, H. Badino, and S. Gehrig. 6D Vision: Fusion of motion and stereo for robust environment perception. In *Pattern Recognition, DAGM Symposium 2005, Vienna*, pages 216–223, 2005. 2

[6] 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

[7] S. Gehrig, H. Badino, and U. Franke. Improving sub-pixel accuracy for long range stereo. *Computer Vision and Image Understanding (CVIU)*, 116(1):16–24, January 2012. 2, 6

[8] S. Gehrig and C. Rabe. Real-time semi-global matching on the CPU. In *ECVW 2010 @ CVPR*, June 2010. 7

[9] 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

[10] R. Gray. *Entropy and information theory*. Springer-Publishing New York, 1990. 5

[11] I. Haller and S. Nedevschi. GPU optimization of the SGM stereo algorithm. In *ICCP*, 2010. 3

[12] A. Hegerath and T. Deselaers. Patch-based object recognition using discriminatively trained Gaussian mixtures. In *BMVC*, 2006. 6

[13] S. Hermann and R. Klette. Iterative SGM for robust driver assistance systems. In *ACCV*, 2012. 6

[14] H. Hirschmueller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *CVPR, San Diego, CA*, volume 2, pages 807–814, June 2005. 1, 2, 3

[15] H. Hirschmueller and S. Gehrig. Stereo matching in the presence of sub-pixel calibration errors. In *CVPR, Miami, FL*, June 2009. 3

[16] H. Hirschmueller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric distortions. *IEEE Transact. Pattern Analysis & Machine Intelligence*, 31(9):1582–1599, 2009. 3

[17] F. Huguet and F. Deverney. A variational method for scene flow estimation from stereo sequences. In *Proceedings of Int. Conference on Computer Vision 07*, 2007. 2

[18] 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*, pages 5198–5203, 2011. 2

[19] S. Larsen, P. Mordohai, M. Pollefeys, and H. Fuchs. Temporally consistent reconstruction from multiple video streams using enhanced belief propagation. In *Proceedings of Int. Conference on Computer Vision 07*, 2007. 2

[20] T. Mueller, J. Rannacher, C. Rabe, and U. Franke. Feature- and depth-supported modified total variation optical flow for 3d motion field estimation in real scenes. In *CVPR*, 2011. 2, 4

[21] H. H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In *ECCV*, pages 139–148, 1990. 2

[22] D. Pfeiffer and U. Franke. Towards a global optimal multi-layer stixel representation of dense 3d data. In *BMVC*, September 2011. 4, 6

[23] D. Pfeiffer, S. Gehrig, and N. Schneider. Exploiting the power of stereo confidences. In *CVPR*, June 2013. 1, 6, 7

[24] C. Rabe, T. Mueller, A. Wedel, and U. Franke. Dense, Robust, and Accurate Motion Field Estimation from Stereo Image Sequences in Real-Time. In *ECCV*, volume 6314, pages 582–595, September 2010. 2, 6

[25] C. Reichardt, D. Orr, I. Davies, A. Criminisi, and N. Dodgson. Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid. In *ECCV*, October 2010. 2

[26] D. Scharstein and R. Szeliski. Middlebury online stereo evaluation, viewed 2013/03/12. http://vision.middlebury.edu/stereo. 1

[27] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, pages 519–528, 2006. 2

[28] F. Stein. Efficient computation of optical flow using the census transform. In *DAGM 2004*, September 2004. 4

[29] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for TV-L_1 optical flow computation. In *Dagstuhl Visual Motion Analysis Workshop*, 2008. 2

[30] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *ECCV*, pages 151–158, 1994. 3