Iterative Semi-Global Matching for Robust Driver Assistance Systems

Simon Hermann and Reinhard Klette

The *.enpeda..* Project, Department of Computer Science
The University of Auckland, New Zealand

Abstract. Semi-global matching (SGM) is a technique of choice for dense stereo estimation in current industrial driver-assistance systems due to its real-time processing capability and its convincing performance. In this paper we introduce iSGM as a new cost integration concept for semi-global matching. In iSGM, accumulated costs are iteratively evaluated and intermediate disparity results serve as input to generate *semi-global distance maps*. This novel data structure supports fast analysis of spatial disparity information and allows for reliable search space reduction in consecutive cost accumulation. As a consequence horizontal costs are stabilized which improves the robustness of the matching result. We demonstrate the superiority of this iterative integration concept against a standard configuration of semi-global matching and compare our results to current state-of-the-art methods on the KITTI Vision Benchmark Suite.

1 Introduction

Semi-global matching (SGM) as proposed by Hirschmüller [8], is used in industrial vision-based driver assistance systems [2] due to its reliable performance. The inherent parallelism of the accumulation strategy supports stereo processing in real-time.

However, a major limiting factor for real-time SGM implementations is the limited memory throughput in current computer hardware. Because SGM integrates aggregated costs along multiple 1-dimensional paths, a large memory block, the *accumulation buffer*, needs to be updated in off-chip memory. Therefore, current literature on real-time SGM [2, 12] proposes to alter the design of the original method by easing the memory requirement to ensure high frame rates for image resolutions of at least 640×480 pixels.

For example, Gehrig et al. [2] propose a design concept for an FPGA implementation of SGM. They calculate a disparity image for a down-scaled stereo pair that serves as a *disparity prior*. A half resolution SGM map only requires $\frac{1}{8}$ of the memory. For full-resolution SGM they restrict the image domain to a specified region-of-interest, and only process the subset $\{0, ..., \frac{d_{max}}{2}\}$ of the *disparity search space* $\mathbb{D} = \{0, \ldots, d_{\max}\} \subset \mathbb{N}$. They generate the final result by replacing prior disparities by disparities from the full-resolution map, if the prior suggests

that a disparity lies within the reduced search space. Otherwise, the prior disparity is taken as the final result. This approach is based on the argument that for objects at close distance, sufficient disparity accuracy can be ensured when computing half-resolution disparity images only. But, as the re-projection error increases quadratically when disparities are smaller and object boundaries far away are often vague due to downscaling, there is a need to calculate disparities at full resolution to minimize uncertainties for distant objects. This approach reduces the memory requirement by 50% for the selected region of interest of the full resolution SGM.

In our previous work [6] we also propose the calculation of a half-resolution disparity prior. But, instead of merging two disparity results at the end, the prior information is used to reduce the search space for stereo matching at full-resolution. Cost values are only processed within a small search interval around the prior disparity. Given a dense disparity prior, the memory requirement for full resolution SGM is reduced to a memory block that is primarily based on the image dimension, and not on the search space limit. The search space itself is adapted pixel-wise, motivated by the objective to accurately calculate disparities regardless of the actual re-projection error.

Although these two SGM methods have been successfully applied for stereo matching, they have a conceptual problem. If prior disparities are estimated incorrectly, it directly affects the final result. This is because either false disparities are replaced (or not replaced), or because the true disparity does not lie within the anticipated search interval. We believe that the major problem is that prior disparities are only considered pixel-wise, ignoring potentially different disparities in the immediate neighborhood.

In our novel approach we avoid fixed pixel-wise prior disparities. Instead of evaluating integrated costs of a half-resolution SGM, we use those costs to initialize the accumulation buffer for the consecutive full-resolution process. By using costs rather than prior disparities, we avoid to constrain 'uncertain disparity estimates' while, at the same time, 'strong disparities' are supported in the consecutive integration process. In order to reliably reduce the search space, we iteratively *evaluate* the accumulation buffer to a disparity image. We perform a *winner-takes-all* (WTA) evaluation during the integration process. To avoid pixel-wise evaluation of the disparity prior, we introduce a novel data structure that allows us to analyse *spatial consistency properties* of a disparity map. Only if a pixel is sufficiently supported by a *semi-global neighborhood*, the search space is reduced.

The concept of search space reduction in stereo matching has been discussed before (see Geiger et al. [4], as one example) and different approaches have been published to perform this task in a reliable way. For example, Gong and Yang [5] proposed an iterative approach where disparities are locked based on a confidence measure of the accumulated cost.

The paper is structured as follows. Section 2 recaps the SGM strategy of [8] and outlines the chosen configuration for the paper. Section 3 introduces our

novel *iterative* SGM (iSGM) approach in detail. Section 4 presents a thorough evaluation and discussion on real-world data. Section 5 concludes.

2 Standard Semi-Global Matching

The SGM concept [8] extends single-line dynamic stereo matching [11] into a multi-line integration strategy. It consists of two steps. First, the costs of pixel correspondences are established for all *disparities* d in the defined search space $\mathbb{D} = \{0, \dots, d_{max}\}$ of non-negative integers. Second, these calculated costs are regularized along scan lines that run across the image domain by employing an accumulative dynamic programming scheme. Accumulated regularization costs from multiple scan lines with different directions are then integrated, and final disparities are selected based on a WTA evaluation.

2.1 Census Cost Function

The costs of pixel correspondences are established based on a *cost function* that defines a similarity value between pixel neighborhoods in the left and right image of a stereo pair. The *census cost function* has been identified [7, 9] to be very descriptive and robust in combination with SGM, especially under strong illumination variations. It is based on the *census transform* as introduced by Zabi and Woodfill [18]. A binary *signature* vector is assigned to each pixel position $p = (i, j)$ of the left and right image. The signature is calculated based on the ordinal characteristic of the intensity $I_p = I_{(i,\ j)}$ of an image I in relation to intensities within a defined neighborhood. This transform is performed once on the left and the right image, prior to cost calculations. Signatures are stored as bit strings in an integer matrix of the same dimensions as the given image. The signature vector is generated as follows:

$$\text{census}_{\text{sig}}(I_{(i,\ j)}) = \Big\{ \Upsilon\big[I_{(i,\ j)} \geq I_{(i+x,\ j+y)}\big] \Big\}_{(x,\ y) \in \mathcal{N}} \tag{1}$$

where $\Upsilon[\cdot]$ returns 1 if true, and 0 otherwise. $\mathcal{N}$ denotes a neighborhood centered at $(0, 0)$. The actual *census cost* is the Hamming distance of two signature vectors and can be calculated very efficiently [15].

2.2 Cost Regularization and Integration

The regularization of correspondence costs is implemented as cost accumulation along oriented *scan lines*, which are 1-dimensional linear paths identified by a direction vector $\mathbf{a}$. The cost $L_{\mathbf{a}}$, defined for a pixel location p and a disparity d, is accumulated between image border and p. Consider the segment $p_0, p_1, \dots, p_n$ of that path, with p_0 on the image border, and $p_n = p$. The cost at p for disparity $d \in \mathbb{D}$, on a scan line defined by $\mathbf{a}$, is recursively defined as follows, for $i = 1, 2, \dots, n$:

$$L_{\mathbf{a}}(p_i, d) = C(p_i, d) + \mathcal{M}_i - \min_{\Delta \in \mathbb{D}} L_{\mathbf{a}}(p_{i-1}, \Delta) \tag{2}$$

Direction **a**	$\longrightarrow$	$\longleftarrow$	$\uparrow$	$\downarrow$	$\nearrow$	$\swarrow$	$\searrow$	$\nwarrow$
Index k	0	1	2	3	4	5	6	7

Table 1. Used indices k of eight directions **a**.

with

$$\mathcal{M}_i = \min \begin{cases} L_{\mathbf{a}}(p_{i-1}, d) \\ L_{\mathbf{a}}(p_{i-1}, d-1) + P_1 \\ L_{\mathbf{a}}(p_{i-1}, d+1) + P_1 \\ \min_{\Delta \in \mathbb{D}} L_{\mathbf{a}}(p_{i-1}, \Delta) + P_2 \end{cases} \tag{3}$$

where $C(p, d)$ is the correspondence cost when matching p for disparity d. Regularization penalties P_1 and P_2 enforce piecewise disparity consistency along a scan line.

SGM uses eight different path directions **a** for scan line accumulation (right, left, up, down, and the four in-between angles). As a core concept of SGM the accumulated costs of all eight scan lines are integrated and the optimum disparity d_{opt} is identified by the minimum integrated cost. We assign unique indices k to directions **a** as in Table 1.

2.3 Penalty Adjustment

Penalties P_1 and $P_2^\star$ are given as external parameters. They implement the Potts model as follows. For a solution d, a constant penalty cost of P_2 is assigned to all disparities $g \neq d$ starting at p. By penalizing all disparities equally, disparity jumps at depth discontinuities are preserved. In order to model smooth transitions of non-fronto parallel surfaces, a smaller penalty $P_1 < P_2$ is assigned to all disparities g that lie within the immediate disparity neighborhood of d, (i.e. $|g - d| = 1$). P_2 is constant for all labels, but it is locally adjusted for each pixel p_i as follows:

$$P_2(p_i) = \max \left\{ \frac{P_2^\star}{|I(p_{i-1}) - I(p_i)|}, P_1 + \delta \right\} \tag{4}$$

where $\delta > 0$. This adjustment links the regularization procedure with the underlying image data since the magnitude of the forward difference in direction **a** scales the penalty at each p_i. The rationale behind this is to improve the performance at depth discontinuities as they are more likely to occur at intensity edges. Another motivation is to reduce the streaking effect, which is inherent to scan line optimizations.

2.4 Sub-pixel Accuracy and Consistency Check

Disparities are calculated with sub-pixel accuracy in an interval $[0, d_{\max}]$ of real numbers using the equiangular interpolation method proposed by Shimizu and Okutomi [14].

To enforce uniqueness of calculated disparities, a second disparity map is calculated to perform a final left-right (L-R) consistency check. A disparity is *valid* if it differs from the corresponding disparities by θ at most. A label inv with a negative value is assigned to a pixel where an invalid disparity has been calculated.

The original SGM strategy [8] applies a median filter on calculated disparities to eliminate outliers and to increase the density of the disparity map. We could not identify a qualitative benefit for our results, and therefore omit the application of a median filter in our implementation.

3 Iterative Semi-Global Matching

The previous section described a specific SGM configuration that follows the integration strategy as introduced by the original SGM algorithm [8]. This section proposes a novel integration strategy.

3.1 Homogeneous Disparity Maps

Consider a disparity map D. The value at a pixel p is either a non-negative disparity $d \in [0, d_{max}] \subset \mathbb{R}$, or the negative label inv indicating an invalid value. In D we change a valid disparity d_p at pixel p into label inv if there is at least one valid disparity d_q at a pixel q being 8-adjacent to p such that $|d_p - d_q| > \gamma$, for a given threshold γ. In other words, if any two 8-adjacent disparities vary by more than γ disparity units then both pixels are invalidated. We refer to the result of this process as being a *homogeneous disparity map* (HDM), homogenized by threshold γ.

3.2 Semi-Global Distance Maps

We introduce a data structure called *semi-global distance map* (SGDM) which is derived from a homogeneous disparity map H. In H, each pixel is either categorized by a valid disparity $d \geq 0$, or by an invalid label inv < 0. These two pixel categories define a binary interpretation of H which is required to generate the SGDM. First, for any of the eight directions $\mathbf{a}$ identified by indices $k = 0, .., 7$, we generate a distance map F_k for the given homogeneous disparity map H as follows:

$$F_k(p_i) = \begin{cases} +1 & \text{if } H(p_{i-1}) < 0 \text{ and } H(p_i) \geq 0 \\ F_k(p_{i-1}) + 1 & \text{if } H(p_{i-1}) \geq 0 \text{ and } H(p_i) \geq 0 \\ -1 & \text{if } H(p_{i-1}) \geq 0 \text{ and } H(p_i) < 0 \\ F_k(p_{i-1}) - 1 & \text{if } H(p_{i-1}) < 0 \text{ and } H(p_i) < 0 \end{cases} \tag{5}$$

where

$$F_k(p_0) = \begin{cases} +1 & \text{if } H(p_0) \geq 0 \\ -1 & \text{if } H(p_0) < 0 \end{cases} \tag{6}$$

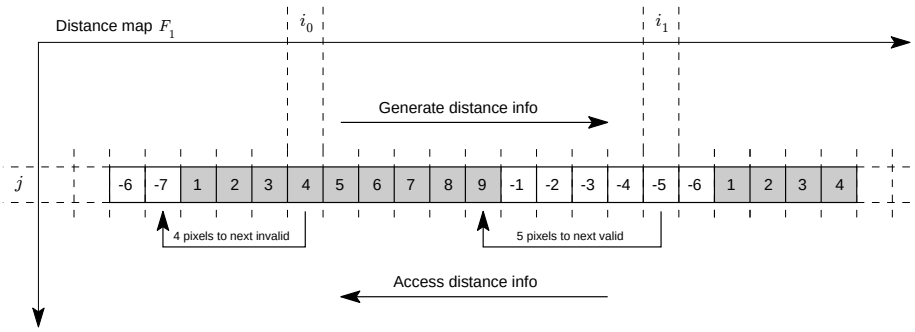

Fig. 1. Example: generation of distance information for F_1.

Recall that p_0 is a pixel at the image border. After generating these distance maps, we relabel them in the following way:

$$F_k = \begin{cases} F_{k+1} & \text{if } k \pmod 2 = 0 \\ F_{k-1} & \text{if } k \pmod 2 = 1 \end{cases} \tag{7}$$

The SGDM is the family of all eight distance maps, and denoted by $\mathcal{F}$. The value $\mathcal{F}_p$ at a pixel p is an 1×8 integer vector

$$\mathcal{F}_p = (f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7)^p \tag{8}$$

that encodes pixel offsets required to access the closest pixel of the opposite pixel category along path k.

Figure 1 shows an example for a single image row of a homogeneous disparity map H. White pixels refer to invalid disparities (label inv in H) while grey pixels indicate a non-negative valid disparity. To generate distance information (in this case for F_1) we either increment a counter when iterating (from left to right) over valid disparities, or decrement a counter when iterating over invalid disparities. The counters are re-initialized with 1 or -1 according to Eqn. (5) whenever the pixel category changes. Although we generate the distance information in direction $k = 0$, we access this information in opposite direction. Therefore, we relabel the distance maps according to Eqn. (7).

Let us consider a standard distance map, which is an image that encodes the distance of an *object pixel* to the closest *edge pixel*. A major advantage of SGDMs is that the distance information at each pixel encodes the actual locations of the eight neighboring pixels of the opposite category. This means that at invalid labels, we get the location of the closest valid disparity along path k, and vice versa. Because we operate on HDMs, we additionally know at valid disparities that any two consecutive disparity values along the path do not vary by more than γ until the next invalid label is reached. These features are exploited later when analyzing intermediate disparity results to reduce the disparity search space.

The generation time for the SGDM is linear in the number of pixels. It is suitable for highly-parallelized processing because directions and individual scan lines are treated independently.

3.3 Putting Things Together

Figure 2 shows a flow-chart of the proposed iterative integration strategy. Rectified stereo pairs are the input of the algorithm. The algorithm starts with

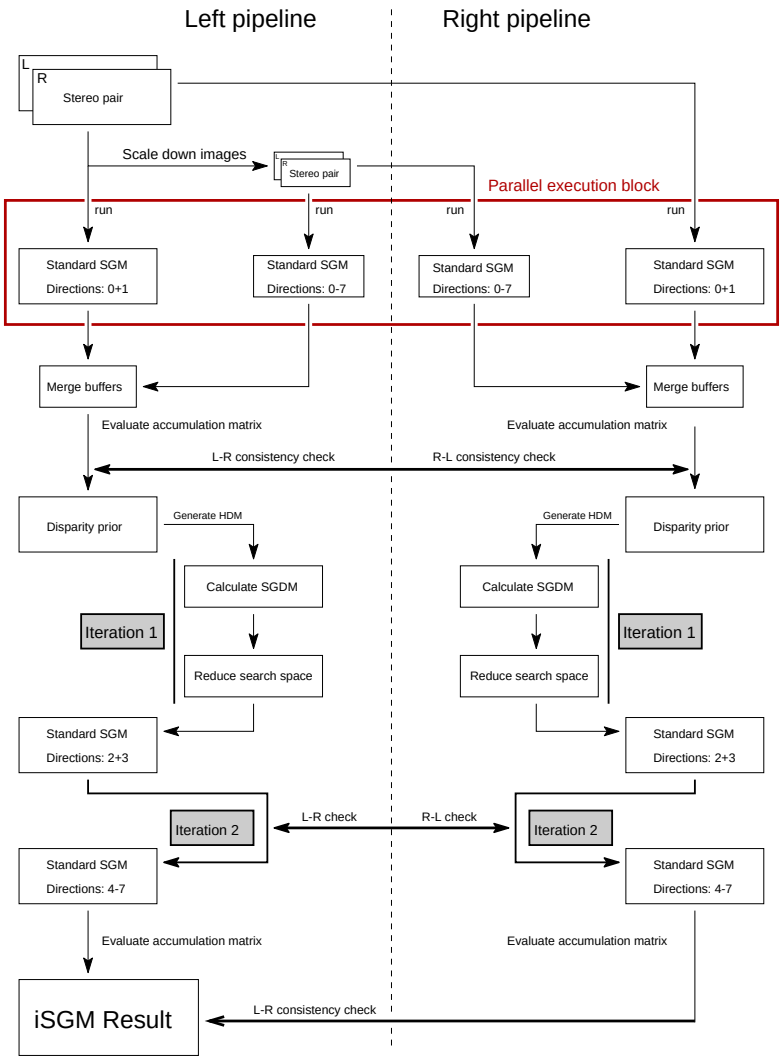

Fig. 2. Flow-chart of the iSGM algorithm.

four parallel SGM processes, denoted by 'Standard SGM' in the chart. They represent standard cost-accumulation as described in Eqn. (2). Directions used for subsequent cost integration are given as additional information in boxes in Fig. 2. Cost calculation is excluded from the chart for better readability. It can either be done on-the-fly (suitable for GPUs, see [1]), or once as a lookup-table prior to cost accumulation.

The outer two of the four SGM processes operate on the original input images, while the inner two on down-scaled versions of these images. The chart is split into a left and a right *processing pipeline* that are executed in parallel threads. Those threads are joined at consistency checks, denoted by bold arrows crossing the dashed middle line. The left pipeline uses the left image as the reference to be matched with the right image. The right pipeline uses images in reverse order.

After these four initial SGM processes are finished, the accumulation buffers of the half-resolution processes contain integrated costs of all eight directions as shown in Table 1. The full-resolution accumulation buffer contains only horizontally accumulated costs (directions 0 and 1). Half- and full-resolution SGM processes are joined by merging costs of each half-resolution accumulation buffer into the corresponding full-resolution accumulation buffer. The impact of half-resolution costs is scaled by a factor η. Both accumulation buffers are then evaluated, and the result defines a disparity map. The execution of both pipelines is mutually dependent in the sense that they are joined by an L-R and R-L consistency check. Disparity images, generated by this check, are used as input to generate disparity priors in order to reduce the search space for subsequent cost-accumulation procedures. This is done for the left and right pipeline, as outlined in the following paragraph.

Search Space Reduction. First we generate a homogeneous disparity map (HDM) H as described in Section 3.1. For H we generate spatial distance information by calculating a semi-global distance map (SGDM). Now, assume that the minimum entry $f_{\min}$ of vector $\mathcal{F}_p$ at pixel p is greater than a given integer distance threshold ϑ. Because we operate on a homogeneous disparity map, we can conclude that, within a $(2 \cdot f_{min}+1) \times (2 \cdot f_{min}+1)$ semi-global neighborhood (defined by directions given in Table 1), all adjacent disparities cannot vary by more than threshold γ.

We assume that, if $f_{min} > \vartheta$ for a sufficiently large ϑ, the likelihood is high that the evaluated cost at p is close to specifying the true disparity. For a valid disparity d_{opt} at pixel p we reduce the search domain as follows:

$$L_{\mathbf{a}}(p_i, d) = C(p_i, d) + \mathcal{M}_i - \min_{\Delta \in \mathbb{D}} L_{\mathbf{a}}(p_{i-1}, \Delta)\Big|_{d=d_{opt}-2}^{d_{opt}+2} \tag{9}$$

By incorporating the semi-global neighborhood we avoid the problem of pixel-wise evaluation of prior disparity maps, as mentioned in Section 1.

Additionally, we leverage the dual representation of the SGDMs to reduce the search space for invalid pixels. Let p be a pixel with label inv. The vector $\mathcal{F}_p$ only contains negative integer values that encode the distance to the next

valid disparity along path k. In consequence, we can access all eight disparity values in constant time, and select the maximum and the minimum to restrict the search space. This is based on the assumption that the true disparity must lie within those limits as these are the eight paths passing through pixel p. We found that this feature is especially useful in the case of homogeneous surfaces.

Note that the dual representation can also be used to define a *semi-global interpolation* for invalid pixels. The weighted mean of the eight valid *semi-global neighbours* define the final value. Weights are defined by the known relative distance values.

The left and right pipelines continue to be processed in parallel with vertical cost accumulation, but now operating on a reduced search space. In order to further reduce the search space, another iteration of an early accumulation buffer evaluation can be performed. Usually, two iterations are sufficient. After that, the diagonal cost-accumulation buffers are evaluated, and one final consistency check is performed to generate the final result.

4 Evaluation and Discussion

This section starts with a discussion that provides insights into iSGM. It highlights its benefits on a qualitative level followed by a quantitative evaluation on a recently released online benchmark dataset. The evaluation shows that iSGM can compete with current state-of-the-art algorithms and demonstrates a comparative performance advantage of iSGM against our standard SGM configuration.

4.1 Inside iSGM

In the following we discuss the novel ideas of iSGM on visualizations of intermediate results that highlight the robustness of this method, especially on challenging data. We therefore selected two frames from the dataset that is provided for the currently running HCI Robust Vision Challenge[1].

The iSGM configuration, as outlined in Section 3.3, employs a coarse-to-fine scheme that scales up costs rather than disparities. This avoids the usage of disparity priors which were previously selected based on a pixel-wise WTA evaluation. In that case, an incorrect cost minimum would lead to a false disparity prior for subsequent processing. In iSGM we initialize the full resolution accumulation buffer with integrated costs from all eight directions of the half resolution SGM. This reduces the impact of horizontal cost accumulation and eases the streaking effect when we evaluate the buffer after horizontal integration. As a consequence, vertical structures which are visible in lower resolution images, are more likely to be preserved and not over-regularized. In other words, it prevents incorrect disparity lock-ins during the first search-space reduction step. See Figure 3, which shows the first frame of the *Blinking Arrow* sequence of the HCI dataset [10]. The left image shows the left input frame with a highlighted region

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

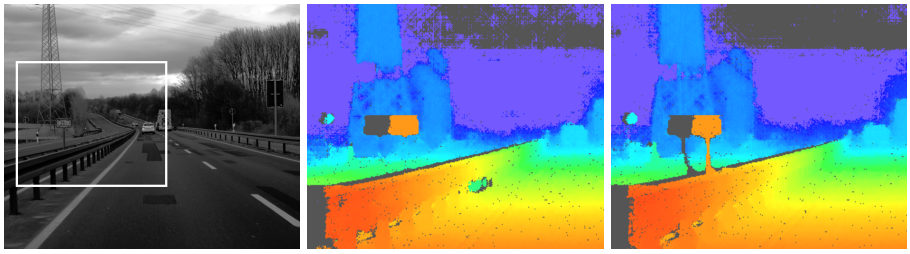

Fig. 3. Frame 0 from the *Blinking Arrow* sequence of the HCI dataset.

of interest. The center and right image show iSGM results for that region. In the center image, we set $\eta = 0$ (i.e. we omit cost initialization). The right image shows the result with cost initialization $\eta = \frac{1}{2}$. All other parameters are identical. We see that the pole-like structure is over-regularized in the center image and therefore fails to be recovered by vertical cost accumulation, because disparities are locked-in by the search space reduction step. Since the initializing costs result from a full standard SGM run, all eight path directions are incorporated on the lower level, such that the horizontal over-regularization is interrupted at crucial regions, which means vertical integration can reconstruct the vertical pole in subsequent processing steps.

However, the path order in which accumulated costs are integrated, is a crucial component of iSGM. To highlight the rationale and potential behind the integration order we present two examples. Figure 4 shows the first stereo pair of the *Rain Blur* sequence, also taken from the HCI dataset. The top-left stereo pair shows an ally road on a rainy day with a wiper blade partially blocking the left part of the right input image. The two top-right images display stereo results of standard SGM and iSGM, respectively. We see that iSGM is clearly outperforming standard SGM in this case. The reason is that we emphasize horizontal cost accumulation. To highlight this effect we run iSGM without cost initialization and display two HDMs for both the *Rain Blur* (bottom left) and for the *Blinking Arrow* (bottom right) sequence. On the left of each resulting pair, we chose horizontal accumulation direction prior to search space reduction. On the right we chose to start with the vertical direction. Disparity values which are invalidated by the HDM process are highlighted by white pixels.

In the *Rain Blur* results, we see that the road is reasonably well recovered in case of horizontal, but totally destroyed in case of vertical accumulation (as it would also be in the diagonal case). We believe that this is because horizontal regularization helps to propagate the underlying weak data cost on a correct disparity level, which does not change significantly in horizontal direction in case of the ground manifold. Because the data cost is 'weak', vertical accumulation can not ensure smooth adaptation of depth levels along the road surface. This is different in case of the *Blinking Arrow* results, where the road is much better recovered in case of vertical accumulation. However, interruptions are clearly visible. We argue that these effects are not sufficient to 'break' the road surface in the case of standard SGM, but they are often visible as 'disparity bumps' in

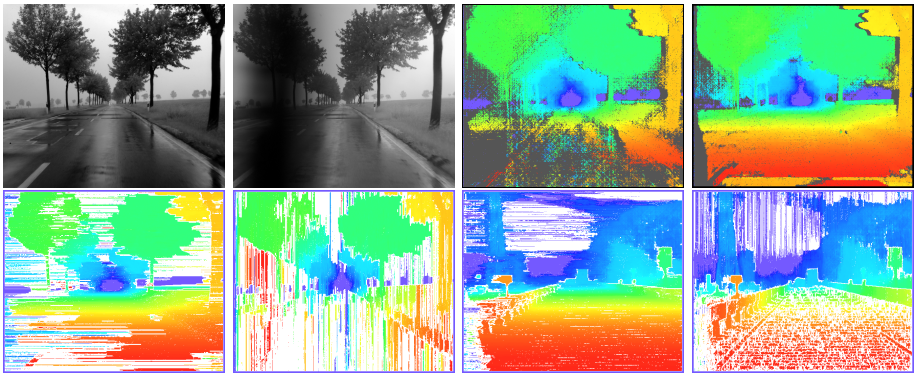

Fig. 4. Frame 0 from the *Rain Blur* sequence taken from the HCI dataset.

the road. We conclude that by evaluating the accumulation buffer after horizontal integration and performing the proposed search space reduction based on SGDMs, correct road disparities are locked-in and are not 'destroyed' by vertical (or non-horizontal) integration.

In extension of this conclusion, we argue that path directions should be 'promoted' based on surface slopes of the given scene geometry. For example, vertical integration should perform best in case of surfaces orthogonal to the road manifold and parallel to the driving direction, such as houses, parked cars, or vertical pole-like structures. We believe that SGDMs can contribute to the solution of this task when directly incorporated into the cost regularization process.

4.2 Evaluation on the KITTI Vision Benchmark Suite

Geiger et al. [3] recently introduced The KITTI Vision Benchmark Suite[2]. It defines a relatively demanding benchmark that currently consists of 195 testing and 194 training stereo pairs with semi-dense ground truth generated by a Velodyne HDL-64E laser range-finder. The images were taken from recorded sequences of real-world driving scenarios. Stereo matching algorithms evaluated on this dataset have to deal with challenging illumination conditions, high image resolution (1240×376 pixels), and large pixel displacements (>128 disparities) due to a 54 cm baseline.

We performed two evaluations. First, we submitted our results for the testing dataset to the online stereo benchmark to compare iSGM with current state-of-the-art algorithms. Second, we used the training dataset that comes with ground truth to perform a comparative evaluation of iSGM against a standard SGM configuration to demonstrate the relative performance gain. We used a fixed parameter set for all following evaluations.

[2] `http://www.cvlibs.net/datasets/kitti/`

KITTI Online Evaluation. At the time of writing, our submission to the KITTI Benchmark Suite ranks second w.r.t the reference evaluation. Reference evaluation is performed on a 100% dense map. Disparities are considered to be 'correct' if they do not deviate by more than 3 disparity units from the ground truth. If a method does not provide 100% disparity density, such as iSGM, a simple background interpolation technique is applied prior to evaluation.

Table 2 shows the performance of the current top-three algorithms. iSGM has only an error difference of 1.11% to the leading belief propagation-based method PCBP proposed by Yamaguchi et al. [16]. Considering that PCBP uses 4 cores for processing, our current implementation runs more than 70 times faster than PCBP. Ranked third is a variational-based method (ITGV) proposed by Ranftl et al. [13].

Rank	Method	Out-Noc	Avg-Noc	Density	Runtime
1	PCBP	4.05 %	0.9 px	100.00 %	300 s
2	iSGM	5.16 %	1.2 px	94.70 %	8 s
3	ITGV	6.31 %	1.3 px	100.00 %	7 s

Table 2. Top three stereo algorithms on the KITTI stereo page on 30 September 2012.

Algorithm Configurations. Our iSGM implementation is C++ based, and the algorithm is executed on an Intel Core2Duo. We utilize hyper-threading to run the left and the right pipeline in separate hardware threads. However, we neither perform any computation on a GPU, nor employ any SSE optimization for data parallelization. The latter is common for SGM implementations. Since accumulated costs are stored in 16 bit integers, there is a significant speed-up potential by processing eight data updates in parallel, utilizing the 128 bit MMX registers. However, our implementation is designed to provide an efficient research tool, and is not optimized for fast execution.

The algorithms are configured as follows. We use $P_1 = 35$ and $P_2^\star = 350$ for an intensity domain of $[0, 255]$, and set $\delta = 1$ to ensure $P_2 > P_1$. The consistency check threshold is set to $\theta = 3$. iSGM introduces three additional parameters. The threshold γ is required to generate the HDM, threshold ϑ defines the required minimum of a semi-global distance vector in order to reduce the search space, and factor η scales the half-resolution accumulation costs before the buffers are merged. For our experiments, we used $\gamma = 2$, $\vartheta = 5$, and $\eta = \frac{1}{2}$. The census cost window was set to 64bit with a star shaped window in case of half resolution and standard SGM. The full resolution data costs in iSGM were of 32 bit length, representing a cross shaped window (7×3 crossed by 3×7 rectangular neighborhoods).

We also evaluated (but did not publish) the result of our standard SGM configuration which was 8.35%. This result deviates only by 0.71% from the OpenCV SGM implementation submitted to KITTI under the name of OCV-SGBM. We therefore assume that we have a reasonable standard SGM configuration available for the following comparative evaluation.

Comparative Evaluation. We evaluated the performance of our chosen standard SGM configuration against our novel iSGM concept. We used the available ground truth of the KITTI training dataset and performed the reference evaluation as described in the previous section.

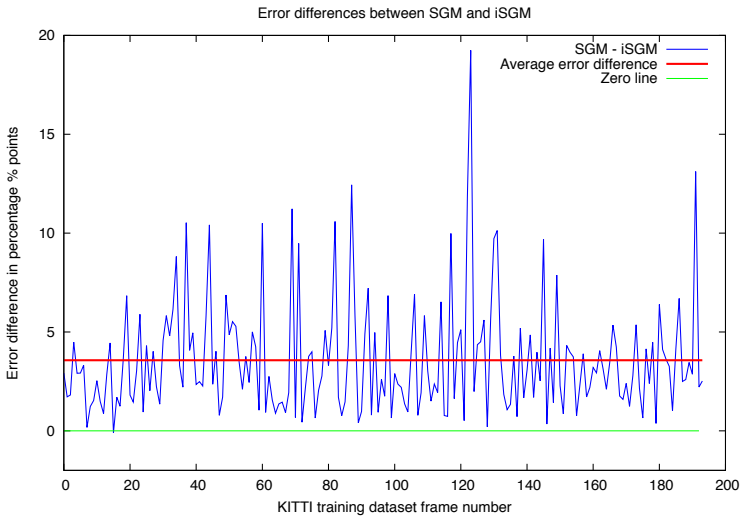

Fig. 5. Error differences (SGM errors minus iSGM errors).

Figure 5 shows the difference (in percentage) between evaluated errors for both methods. A positive value means that iSGM outperforms standard SGM. With exception of frame 15 of the KITTI testing dataset (error difference is -0.098%), iSGM is always outperforming our standard SGM configuration. In average, the improvement equals 3.57%. We conclude that this performance gain must entirely be due to the novel integration concept, as both methods use identical regularization parameters and procedures as described in Section 2.

5 Conclusions

In this paper we presented iSGM, a new approach to improve the robustness of stereo processing in context of semi-global matching. iSGM translates the conceptual parallel integration strategy into an iterative scheme. We introduce a novel data structure, the semi-global distance map, which can be employed for effective spatial evaluation. We use SGDMs to iteratively reduce the search space by locking reliable disparities from a pre-evaluated disparity prior. This promotes horizontally accumulated costs and as a consequence stabilizes road surfaces especially on challenging stereo data. iSGM currently ranks second on the KITTI Vision Benchmark Suite.

References

1. Ernst, I., Hirschmüller, H.: Mutual information based semi-global stereo matching on the GPU. In Proc. Int. Symp. Advances Visual Computing (ISVC) (2008) 228–239
2. Gehrig, S. K., Eberli, F., Meyer, T.: A real-time low-power stereo vision engine using semi-global matching. In Proc. Int. Conf. Computer Vision Systems (ICVS), LNCS 5815 (2009) 134–143
3. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI Vision Benchmark Suite. In Proc. Computer Vision Pattern Recognition (CVPR) (2012)
4. Geiger, A., Roser, M., Urtasun, R.: Efficient large-scale stereo matching. In Proc. Asian Conference on Computer Vision (ACCV), LNCS 6492 (2010) 25–38
5. Gong, M., Yang, Y.-H.: Fast stereo matching using reliability-based dynamic programming and consistency constraints. In Proc. Int. Conf. Computer Vision (ICCV), **1** (2003) 610–617
6. Hermann, S., Klette, R.: Evaluation of a new coarse-to-fine strategy for fast semi-global stereo matching. In Proc. Advances in Image and Video Technology (PSIVT), LNCS 7087 (2012) 395–406
7. Hermann, S., Morales, S., Vaudrey, T., Klette, R.: Illumination invariant cost functions in semi-global matching. In Proc. Computer Vision Vehicle Technology: From Earth to Mars (CVVT:E2M), ACCV workshop, LNCS 6469 (2010) 245–254
8. Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In Proc. IEEE Int. Conf. Computer Vision Pattern Recognition (CVPR), **2** (2005) 807–814
9. Hirschmüller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans. Pattern Analysis Machine Intelligence, **31** (2009) 1582–1599
10. Meister, S., Jähne, B., Kondermann, D.: Outdoor stereo camera system for the generation of real-world benchmark data sets. Optical Engineering **51** (2012) paper 021107, 6 pages
11. Ohta, Y., Kanade, T.: Stereo by two-level dynamic programming. In Proc. Int. Joint Conf. Artificial Intelligence (IJCAI) **2** (1985) 1120–1126
12. Pantillie, C., Nedevschi, S.: SORT-SGM: Subpixel optimized real-time semiglobal matching for intelligent vehicles. IEEE Trans. Vehicular Technology **61** (2012) 369–376
13. Ranftl, R., Gehrig, S., Pock, T., Bischof, H.: Pushing the limits of stereo using variational stereo estimation. In Proc. IEEE Intelligent Vehicles Symposium (IV), to appear (2012)
14. Shimizu, M., Okutomi, M.: An analysis of subpixel estimation error on area-based image matching. In Proc. IEEE Conf. Digital Signal Processing (DSP) **2** (2002) 1239–1242, .
15. Warren, H.S.: Hacker's Delight. Pages 65–72, Addison-Wesley Longman, New York (2002)
16. Yamaguchi, K., Hazan, T., McAllester, D., Urtasun, R.: Continuous Markov random fields for robust stereo estimation. In arXiv:1204.1393v1, (2012)
17. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime TV-L1 optical flow. In Proc. Pattern Recognition (DAGM), LNCS 4713 (2007) 214–223
18. Zabih, R., Woodfill, J.: Non-parametric local transform for computing visual correspondence. In Proc. Europ. Conf. Computer Vision (ECCV) **2** (1994) 151–158