

Computationally efficient dense moving object detection based on reduced space disparity estimation

Goran Popović, Antea Hadviger, Ivan Marković,
Ivan Petrović*

* *University of Zagreb Faculty of Electrical Engineering and Computing, Laboratory for Autonomous Systems and Mobile Robotics, Zagreb, Croatia (e-mail: name.surname@fer.hr).*

Abstract: Computationally efficient moving object detection and depth estimation from a stereo camera is an extremely useful tool for many computer vision applications, including robotics and autonomous driving. In this paper we show how moving objects can be densely detected by estimating disparity using an algorithm that improves complexity and accuracy of stereo matching by relying on information from previous frames. The main idea behind this approach is that by using the ego-motion estimation and the disparity map of the previous frame, we can set a prior base that enables us to reduce the complexity of the current frame disparity estimation, subsequently also detecting moving objects in the scene. For each pixel we run a Kalman filter that recursively fuses the disparity prediction and reduced space semi-global matching (SGM) measurements. The proposed algorithm has been implemented and optimized using streaming single instruction multiple data instruction set and multi-threading. Furthermore, in order to estimate the process and measurement noise as reliably as possible, we conduct extensive experiments on the KITTI suite using the ground truth obtained by the 3D laser range sensor. Concerning disparity estimation, compared to the OpenCV SGM implementation, the proposed method yields improvement on the KITTI dataset sequences in terms of both speed and accuracy.

Keywords: stereo vision, dense disparity estimation, moving object detection, semi-global matching

1. INTRODUCTION

Depth estimation plays an important role in various autonomous mobile robot environment inference applications; such as simultaneous localization and mapping, navigation—particularly obstacle detection and avoidance. Even though depth can be estimated using a variety of sensors, cameras are still a very popular choice due to their low price, high availability, and low weight, especially when payload is a critical factor as in small unmanned aerial vehicles. The depth, or disparity maps (DMs), can be computed from the intensity images obtained by either monocular or stereo cameras. In the analytical domain, general categorization splits the disparity computation methods into two groups: local methods and global methods. Local methods are faster and prone to discontinuities as they consider only a neighborhood of a pixel to determine its depth. On the other hand, global methods typically provide continuous DMs due to optimizing a global cost function, including the smoothness terms, making the

computation more complex. Another class of currently promising approaches are deep-learning based methods, where the DM estimation is learned from a sequence of animated images with readily available ground truth. However, in the present paper we focus on the computational efficiency of an analytical approach, and do not discuss further learning based methods in the present paper.

Hirschmüller (2008) introduced an approach that reduces the complexity of global methods, but still keeps the smoothness of the obtained DM. The proposed method, dubbed semi-global matching (SGM), approximates the global cost function by recursively aggregating costs in several directions across the image, making it faster to compute than global methods, as well as discontinuity aware. SGM is one of the state-of-the-art algorithms for DM estimation and still nowadays researchers try to modify it in order to make it faster and more accurate. One approach is to optimize the hardware architecture for fast disparity computation, Li et al. (2017) and Hofmann et al. (2016). On the other hand, faster execution can be achieved by reducing algorithm’s complexity or by executing more tasks simultaneously. Most common approaches to complexity reductions are achieved by reducing the image resolution (Gehrig and Rabe, 2010), reducing the number of accumulation paths (Hermann et al., 2011), and reducing the disparity range (Fuček et al., 2017). Apart from reduc-

* This work has been supported by the European Unions Horizon 2020 research and innovation programme under grant agreement No. 688117 (SafeLog) and by the Croatian Science Foundation under contract No. I-2406-2016. This work has also been supported by the FLAG-ERA JTC 2016 and the Ministry of Science and Education of the Republic of Croatia under the project Rethinking Robotics for the Robot Companion of the future (RoboCom++).

ing the algorithm’s complexity, execution time can also be improved with studious implementation. Spangenberg et al. (2014) proposed exploiting the CPU architecture by using SIMD instructions and multi-threading. Therein the authors modified the original SGM algorithm by compressing the disparity space (subsampling) and separating the image into horizontal stripes to enable multi-threading in the step of cost accumulation. Newer approaches improve the computation accuracy with neural networks as in Seki et al. (2017) where the authors use learned penalties for disparity jumps.

Since DM computation is often computed on a sequence of images, it is natural to use the information computed in previous steps to help improve computation in the following ones, in terms of both speed and accuracy. Utilization of recursive filters, such as the Kalman filter, has been described in the works of Matthies et al. (1989), Agrawal et al. (2005), Morales and Klette (2013). Work of Matthies et al. (1989) introduces a pixel based approach meaning that Kalman filtering is performed on all the pixels in the image, and not just on distinct points set. In Morales and Klette (2013) authors integrated and compared several DM computation algorithms within a Kalman filtering framework. One of the compared DM algorithms was SGM, which served as a measurement generator for the Kalman filter. Since Kalman filtering yields prediction of the system’s state, consequently it also offers statistical ground for detecting moving objects in dynamic scenes. In the work of Agrawal et al. (2005) authors used a similar approach to track a moving person. Therein, the estimated motion of the camera was used to predict the disparity image, which was then compared with the actual measurement and the areas with low similarity value were grouped into blobs. Blobs were then tracked with the Kalman filter, thus tracking the moving people.

In the present paper we draw upon our earlier work of Fućek et al. (2017) and we fuse the discussed ideas and approaches, in order to produce a computationally efficient algorithm capable of detecting moving object using the reduced disparity search space. We use pixel-wise Kalman filters to estimate the disparity on reduced search space and densely detect parts of the image which contain moving objects. Thereafter, pixels are grouped in order to get a single and more robust detection indicating moving objects. The implemented algorithm is evaluated on the KITTI dataset (Geiger et al., 2013) and compared with OpenCV’s implementation of SGM in terms of execution speed and accuracy. Furthermore, we use the disparity ground truth from the 3D laser range sensor of the KITTI sequences, to estimate the process noise, i.e., the uncertainty of the ego-motion, and measurement noise, i.e., the uncertainty of the SGM-based disparity estimation.

The rest of the paper is organized as follows. First, we describe the steps of the reduced search space SGM algorithm and its efficient implementation in Section 2. In Section 3 we describe how we estimate the process and measurement noise of the Kalman filter and in Section 4 we show how the enhancement of the algorithm can detect the movement of objects in the scene. Experimental evaluation and comparison with the OpenCV implementation of SGM is given in Section 5, while Section 6 concludes the paper.

2. REDUCED SPACE SEMI-GLOBAL MATCHING

We have modified the general framework of SGM in order to enable reduction of the disparity search space. Apart from the intensity image pair, SGM also inputs two values for each pixel that define the range for disparity search. The algorithm has been studiously implemented for efficiency using SSE instruction set and multi-threading. In the sequel we describe the implemented algorithm in details.

2.1 Matching cost calculation

The first step of all stereo matching algorithms is the computation of the matching cost that measures the similarity between the base image patches and potentially matching ones in the matching image. Since the KITTI dataset consists of outdoor environment images with poorly controlled lightning, it is necessary to choose a matching cost method invariant to changes in brightness. The implemented algorithm uses the census-based matching cost, proven to perform well in the described conditions (Hirschmüller and Scharstein, 2009). Census transform describes a pixel using the relative intensity of its neighboring pixels inside an $n \times m$ window. The final cost of matching a pair of pixels is calculated as the Hamming distance between the two obtained census transforms.

For real-time applications, window size of 5×5 pixels yields good performance in terms of both speed and accuracy (Gehrig and Rabe, 2010). The matching cost computation step is performed $height \cdot width \cdot number\ of\ disparities$ times per frame, making it one of the most time-consuming parts of the algorithm. Census transforms of all the pixels are calculated using the SSE instruction set, allowing processing of 16 pixels of an 8 bit image at once using the 128-bit SSE registers. However, due to additional instructions necessary to correctly align the data, the speedup is lower than 16 times. The final matching cost is calculated using the `xor` operation supported by SSE, followed by a population count. Both census transform and cost computation are parallelized by using multiple threads for corresponding horizontal stripes of the image as the calculation is completely independent for each pixel.

2.2 Cost aggregation

Semi-global matching propagates information throughout the image by recursively aggregating matching costs and discontinuity penalties along several one-dimensional paths using the following equation:

$$L_r(\mathbf{p}, d) = C(\mathbf{p}, d) + \min \begin{cases} L_r(\mathbf{p} - \mathbf{r}, d) \\ L_r(\mathbf{p} - \mathbf{r}, d \pm 1) + P_1 \\ \min_i L_r(\mathbf{p} - \mathbf{r}, d + i) + P_2 \end{cases} \quad (1)$$

Path accumulation is the most time-critical part of the algorithm, its complexity being $\mathcal{O}(number\ of\ paths \cdot height \cdot width \cdot number\ of\ disparities)$. Usually, the number of individual paths is 4, 8 or 16, which directly affects execution time. In the conducted experiments, presented in Section 5, we have noticed that increasing the number of

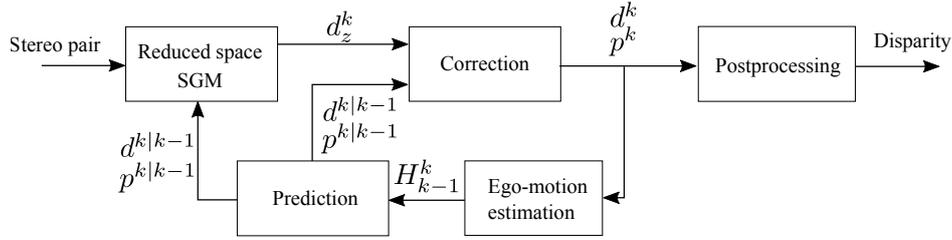


Fig. 1. Block diagram of proposed SGM algorithm.

paths above 4 did not justify the increased execution time, but on the contrary yielded similar results in accuracy. The implementation itself is done using AVX2 instructions, operating on 256-bit vectors. Since accumulated loss can fit into 16-bit integers, it is possible to process 16 disparity values at a time. Costs of individual paths are aggregated into the final loss in the first pass going from the top left to the bottom right, and in the second pass going from the bottom right to the top left. Since this information needs to be propagated throughout the image, it is not possible to parallelize this step on the thread level as in the case of matching cost calculation.

2.3 Disparity computation and refinement

Disparity values are obtained using the simple *winner-takes-all* approach, meaning that the chosen disparity for each pixel is the one with the minimum accumulated loss across all paths. Parallelization in this step is done using SIMD instructions and by dividing the image into horizontal stripes on thread level. Despite the smoothing constraints, noise can be present in the final disparity map. In postprocessing, a block median filter of 3×3 pixels is used to reduce noise, fill the values of possibly invalidated disparities, and additionally improve smoothness. It is worth noting that the median filter is outside of the Kalman loop (Figure 1), thus not affecting the recursive filtering process. However, the downside of median blurring is that it can negatively affect original correctly calculated disparity values.

2.4 Disparity estimation via Kalman filtering

As previously described in the work of Fućek et al. (2017), Kalman filter is used to estimate the disparity and its variance. For completeness, herein we provide the equations of the filtering recursion. The prediction step estimates the coordinates and disparity of all pixels from the frame $k-1$ in the next frame k based on the camera ego-motion. The filter is applied directly in the disparity state-space as suggested by Demirdjian and Darrell (2001). In such a state-space representation, the prediction step of the Kalman filter is given by the following equation:

$$\begin{bmatrix} \omega^{k|k-1} \\ 1 \end{bmatrix} = H_{k-1}^k \begin{bmatrix} \omega^{k-1} \\ 1 \end{bmatrix}, \quad (2)$$

where $\omega = [x \ y \ d]^\top$ are the coordinates in the disparity space and H_{k-1}^k is the 4×4 transformation matrix.

For the computation of disparity variance, we used a motion model presented in Demirdjian and Darrell (2001) where the model is defined as the following ratio of the predicted and current disparity:

$$\Phi^{k-1} = d^{k|k-1} / d^{k-1}. \quad (3)$$

The variance of the prediction is then computed as:

$$p^{k|k-1} = (\Phi^{k-1})^2 p^{k-1} + q^{k-1}, \quad (4)$$

where q^{k-1} is the process noise.

Incorrect disparity predictions are expected near object edges, i.e. disparity discontinuities, due to possible errors in ego-motion estimation. To allow correction of the introduced errors in the stereo matching phase, we reject all the predicted disparities near disparity discontinuities and reset their variance to cover the whole disparity space.

Camera's forward movements introduce holes in the disparity prediction which is known as the *zooming effect*. To cancel it out, we define a new value of disparity and variance for each pixel that has either 2 vertical or 2 horizontal neighbors with known disparity. In that case pixel's disparity and variance are defined as the average of its neighbors. Although in (2) the vector ω has three variables, the steps of Kalman filter are computed for the disparity only, while for the first two coordinates we do not consider their statistical properties.

After the prediction step, the predicted disparity mean and variance are forwarded to SGM to set the limits on the search space to $(d^{k|k-1} \pm p^{k|k-1})$ in the cost aggregation step, corresponding to the possible values of i in the equation (1). Therefore, the search space is reduced proportionally to the uncertainty of the predicted disparity. Given the measured disparity that is then obtained from SGM, it is possible to perform the correction step:

$$d^k = d^{k|k-1} + K_k (d_z^k - d^{k|k-1}) \quad (5)$$

$$p^k = (1 - K_k)^2 p^{k|k-1} + K_k^2 r^k \quad (6)$$

where K_k is the Kalman gain.

Figure 1 shows the block diagram of the overall algorithm. Parameters $d^{k|k-1}$ and $p^{k|k-1}$, which define the search limits, are initially set to cover the whole search space, effectively turning our approach into the conventional SGM. It should also be pointed out that the current algorithm can also work with other recursive estimators that can provide the parameters to determine the search space.

3. PROCESS AND MEASUREMENT NOISE ESTIMATION

In Fućek et al. (2017), the process noise was heuristically determined while the measurement noise was estimated from the cost function values of the neighboring disparities. Since the reduction of disparity space search depends

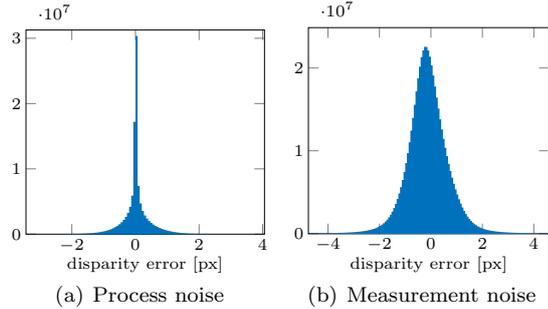


Fig. 2. Distribution of error expressed in pixels. (a) process noise of ego-motion estimation and (b) measurement noise of proposed SGM implementation.

on the estimated covariance, which in turn is also a function of the aforementioned noises, it is important to know them as precisely as possible. Therefore, we estimated the process and measurement noise from the ground truth data of the KITTI dataset that is obtained from the measurements of the 3D laser range sensor. We extracted scenes that did not contain moving objects from 9 different sequences, which summed up to 7000 images in total.

3.1 Process noise

The source of the process noise is the uncertainty of ego-motion estimation. To single out the error occurring due to ego-motion estimation, we used only the static parts of the sequences, without moving objects, where all the motion in the image comes from the motion of the camera. Ego-motion was computed using the open source `libviso2` library (Geiger et al., 2011). Each measurement was extracted using the intensity images of the stereo camera and depth information obtained by the 3D laser range sensor. The ego-motion algorithm estimated the transformation matrix using two consecutive stereo image pairs from the frames $k - 1$ and k . With the computed transformation matrix, we transformed the ground truth from $k - 1$ to estimate the ground truth in k . We assumed that the difference between the estimated disparity and ground truth in k comes from the errors in the transformation matrix, and try to measure the variance of that error in the disparity space. Figure 2(a) shows the histogram of the disparity error in pixels after applying estimated ego-motion, from which we can see that more than 99% errors are within ± 1 pixel. Figure 3 shows the result of error accumulated over two consecutive sequences.

3.2 Measurement noise

To estimate the measurement noise, we measured the error introduced by our implementation of the SGM algorithm. Again, as in the case of process noise estimation, only scenes with static objects are used. We used left and right intensity images and the 3D laser range sensor measurements in the single moment k . Even though the measurements are conducted on image sequences for each image pair, the disparity is computed without reducing the disparity search space. In other words, each image pair was processed as if it appeared at the initialization of the Kalman filter. This way we avoid possible errors due to the reduced space and measure only the uncertainty of

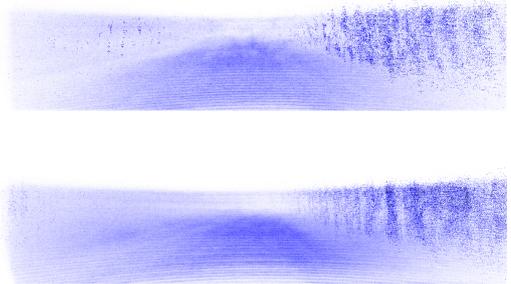


Fig. 3. Accumulated absolute error of 2 different sequences. Intensity of blue color is proportional to the error.

the SGM computation process. The computed disparity between the left and right intensity image is compared to the measurements from the laser sensor and the difference is used for noise estimation. Figure 2(b) shows the measurements of error values on one of the image sequences.

4. MOVING OBJECT DETECTION

The estimated ego-motion can be used to predict correctly the disparity in the future frame for predominantly static scenes; naturally, up to the point of the estimated relative displacement accuracy and the amount of introduced novel structure in the scene. Since in the present paper we are not tracking dynamic objects through time, their displacement as well cannot be accounted for; therefore, the predicted and measured disparity maps will differ in the areas where moving objects are present. Hence, moving objects are detected in the disparity domain without the need to take the intensity into account. The goal is to obtain the bounding boxes surrounding moving objects by detecting the areas that contain high number of pixels whose predicted disparity differs from the measured one.

Moving objects are expected to be located in the parts of the image where the sum of the corresponding values contained in the submatrix containing absolute differences between the predicted and the measured disparities is high relative to its area. The sum of all elements of a submatrix can be calculated in constant time using the *inclusion-exclusion principle* with additional preprocessing. First, small fragments of the image containing high amounts of disparity differences are detected using the sliding window approach. Window sizes used in this step range from 20×20 to 50×75 pixels. Sliding windows are moved only through the bottom two thirds of the image, thus avoiding excess computation as the upper third of the images in the KITTI dataset does not contain moving objects. Due to using the aforementioned window sizes, detected areas of movement intersect and their sizes do not correspond to the exact sizes of the actual moving objects. Those bounding boxes need to be processed further to obtain the wanted result, one bounding box per object, and they are grouped using a greedy approach. In each iteration, a pair of boxes whose area of intersection relative to the area of their union is the highest is replaced by the smallest possible bounding box containing both of them. As the final step, the bounding boxes of too small size are discarded, as they are unlikely to contain moving objects detectable by using the described method.

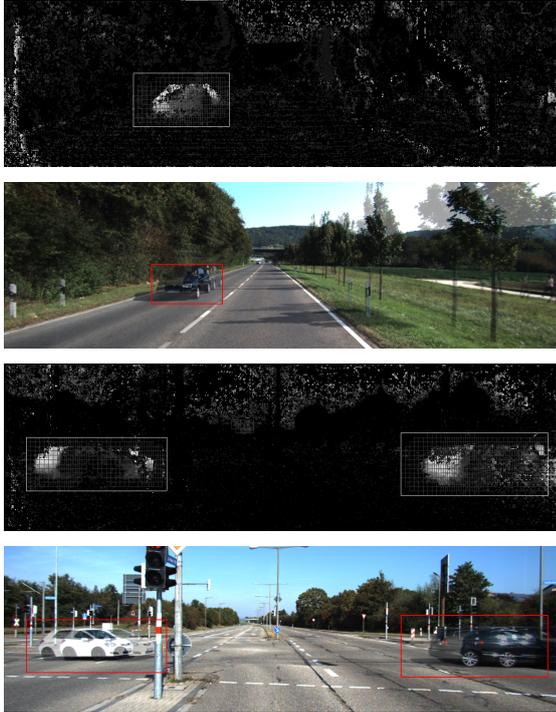


Fig. 4. Moving object detection on two consecutive frames in scenes 183 and 45 from the *data scene flow* set of images. The first and the third image show the difference between the predicted and the matched disparity maps for the respective scenes.

Moving objects in the KITTI dataset are mostly cars, cyclists and pedestrians. With respect to their pattern of movement, they can be divided into two ‘base’ groups: objects moving directly towards and away from the camera and objects moving perpendicular to the direction of camera movement. When the moving object displacement is small, detecting motion for the former group can be challenging since it occurs directly along the projection line of the cameras, while for the latter group (especially large objects) it is possible that there will be no difference in the predicted and the measured disparity in the central area of the object as it does not change the depth of the scene. Nevertheless, these are border cases and throughout the frames most objects exhibit combination of both motions. In the top two images of Fig. 4 we can see a snapshot of detecting an incoming car. It is important to point out that the presented method detects motion through two consecutive frames, as illustrated in the images in Fig. 4, thus making the bounding boxes larger than the actual moving objects in some cases.

5. EVALUATION AND RESULTS

We evaluate the proposed algorithm on the sequences of the KITTI dataset. Due to specific requirements of our algorithm, it was not possible to evaluate it on any of the official benchmarks. The closest benchmarks are the *scene flow* benchmark and the *depth* benchmark. For the evaluation on the *scene flow* benchmark, we would also need the optical flow since the disparity ground truth is given only for the first frame. On the other hand, the *depth* benchmark only has one image per scene and we would not

Table 1. Parameters set in the evaluation

Parameters	OpenCV	proposed
P_1	26	6
P_2	470	65
SADWindowSize	3×3	-
CensusWindowSize	-	5×5
Loss acc. paths	8	4 & 8
Max. disp.	128	128

be able to show how the algorithm works in the reduced space. Thus we used the raw data sequences that provide the ground truth (i.e. laser measurements) for multiple images per scene and for each moment ground truth was in the reference frame of the left image. The implementation of the proposed algorithm is compared to the open source SGM implementation from the OpenCV library.

Before the evaluation, we used parts of the scenes without moving objects for tuning the parameters of the algorithms, namely, the the smoothness factors P_1 and P_2 . The parameters of the proposed algorithm were determined without reducing the disparity search space and without the Kalman filter. The parametrization of OpenCV’s SGM implementation was also conducted on the same sequences, and we tuned the parameters P_1 , P_2 and the size of the window for block matching SADWindowSize, while other parameters were set to default values. The parameters used in the evaluation are shown in Table 1.

Evaluation was conducted on 7 scenes which were not used for the parameter tuning and noise estimation. The results for each sequence are given in Table 2. We can see that in all of the sequences, the average time needed to compute the disparity is smaller for the proposed algorithm than for the OpenCV’s implementation. It should be noted that the proposed algorithm, besides computing disparity using reduced search space, also runs disparity ego-motion based prediction and pixel-wise Kalman filtering. In terms of accuracy, our implementation had on average fewer outliers in 6 out of 7 sequences. In the problematic sequence, Figure 5, the parts of the scene with a lot of moving objects have the highest impact on the error.

Since the number of accumulation paths in our implementation is a variable, in Table 2 we also provide the execution time and accuracy for our implementation when only 4 accumulation paths are used. As expected, the execution time is slightly shorter, but contrary to our expectation the accuracy is better when nondiagonal paths are used. Authors in Spangenberg et al. (2013) explain that due to environment’s structure, horizontal and vertical accumulation paths are more suitable for disparity computation. By computing the disparity with 4 diagonal paths we noticed an increased number of outliers as shown in Table 2. This explains better performance of nondiagonal 4-path implementation since the 8-path implementation takes also the errors of diagonal paths.

6. CONCLUSION

In this paper we have presented how semi-global matching can be improved in terms of both speed and accuracy

¹ Full names of the sequences are: 2011_09_26_drive_{0019, 0039, 0051, 0061, 0084, 0096, 0117}_sync

Table 2. OpenCV and proposed implementation comparison. Outliers are defined as on the KITTI benchmark (absolute threshold of 3 pixels and relative threshold of 5%). Diagonal and nondiagonal columns show the results computed using only 4 of 8 accumulations paths.

Sequence	Time per image [s]			Outliers [%]				No. of images
	OpenCV		proposed	OpenCV		proposed		
	8-path	4-path	8-path	8-path	4-path nondiag.	diag.	8-path	
¹ 19	0.25	0.17	0.19	1.07	0.58	2.11	0.83	395
¹ 39	0.26	0.17	0.19	1.33	1.09	1.85	0.97	384
¹ 51	0.28	0.16	0.18	1.72	2.41	4.49	2.72	427
¹ 61	0.28	0.18	0.19	4.64	1.96	3.74	2.11	691
¹ 84	0.26	0.16	0.18	1.27	0.75	1.70	0.76	372
¹ 96	0.26	0.17	0.19	2.38	1.35	4.18	1.65	464
¹ 117	0.26	0.17	0.19	2.93	1.54	3.88	1.75	649

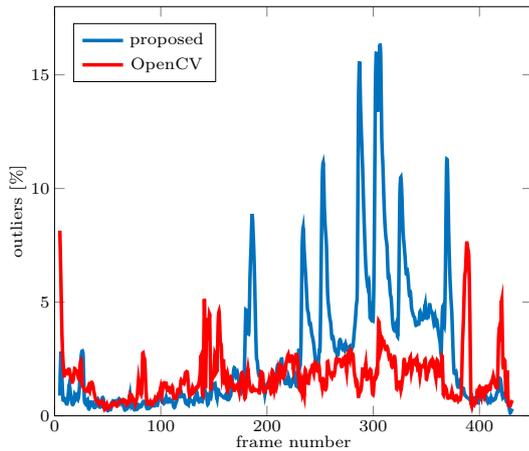


Fig. 5. Comparison of outliers computed with OpenCV SGM and the proposed approach in the sequence 0051¹. The peaks of the blue curve correspond to the frames with moving cars.

by using information from the previous frames, while also detecting moving objects in the scenes. By using the ego-motion estimation and the disparity map of the previous frame, we can predict the disparity of the current frame and use it to reduce the complexity of stereo matching. We use the Kalman filter to fuse the predicted disparity map and the reduced space semi-global matching measurement. The predicted and the measured disparity will differ in the areas containing moving objects since their location cannot be accounted for using only this method. Thus, we show how moving objects can be detected in the disparity domain without the need to consider the intensity images. Furthermore, we estimate the process and measurement noise by conducting extensive experiments on the KITTI suite, on which we also validated our implementation of the reduced space semi-global matching. Compared to the OpenCV SGM implementation, our method yields better results on the KITTI suite in terms of both speed and disparity accuracy.

REFERENCES

Agrawal, M., Konolige, K., and Iocchi, L. (2005). Real-time detection of independent motion using stereo. In *IEEE Workshops on Application of Computer Vision*.
Demirdjian, D. and Darrell, T. (2001). Motion estimation from disparity images. In *IEEE International Conference on Computer Vision, 2001*.

Fuček, L., Marković, I., Cvišić, I., and Petrović, I. (2017). Dense disparity estimation in ego-motion reduced search space. *IFAC World Congress*.
Gehrig, S.K. and Rabe, C. (2010). Real-time semi-global matching on the cpu. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*.
Geiger, A., Lenz, P., Stiller, C., and Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*.
Geiger, A., Ziegler, J., and Stiller, C. (2011). Stereoscan: Dense 3d reconstruction in real-time. In *IEEE Intelligent Vehicles Symposium (IV)*.
Hermann, S., Morales, S., and Klette, R. (2011). Half-resolution semi-global stereo matching. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE.
Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
Hirschmüller, H. and Scharstein, D. (2009). Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
Hofmann, J., Korinth, J., and Koch, A. (2016). A scalable latency-insensitive architecture for FPGA-Accelerated semi-global matching in stereo vision applications. *2016 International Conference on Reconfigurable Computing and FPGAs*.
Li, Y., Yang, C., Zhong, W., Li, Z., and Chen, S. (2017). High throughput hardware architecture for accurate semi-global matching. *Proceedings of the Asia and South Pacific Design Automation Conference*.
Matthies, L., Kanade, T., and Szeliski, R. (1989). Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*.
Morales, S. and Klette, R. (2013). Kalman-filter based spatio-temporal disparity integration. *Pattern Recognition Letters*.
Seki, A., Pollefeys, M., Corporation, T., Zürich, E.T., and Microsoft (2017). SGM-Nets: Semi-global matching with neural networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition*.
Spangenberg, R., Langner, T., and Adfeldt, S. (2014). Large Scale Semi-Global Matching on the CPU. In *IEEE Intelligent Vehicles Symposium (IV)*.
Spangenberg, R., Langner, T., and Rojas, R. (2013). Weighted semi-global matching and center-symmetric census transform for robust driver assistance. In *International Conference on Computer Analysis of Images and Patterns*.