Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Block-Matching Stereo with Relaxed Fronto-Parallel Assumption

Nils Einecke, Julian Eggert

2014

Preprint:

This is an accepted article published in Intelligent Vehicles Symposium. The final authenticated version is available online at: https://doi.org/[DOI not available]

2014 IEEE Intelligent Vehicles Symposium (IV)
June 8-11, 2014. Dearborn, Michigan, USA

Block-Matching Stereo with Relaxed Fronto-Parallel Assumption

Nils Einecke[1] and Julian Eggert[1]

Abstract— **In this paper, we present a new scheme for block-matching stereo. The main intention is to relax the inherent assumption that within a block the disparities are constant, because this assumption is often violated. Instead of using the matching cost of one disparity within a matching block, the best matching for several disparities are first selected for each pixel and then these best matches are combined to the final block-matching value. Results on the KITTI benchmark show that this scheme increases the performance of block-matching stereo especially for large matching windows, however, there is also a significant increase for smaller block sizes. Furthermore, we show that a straightforward combination with the appearance-aligned block-matching stereo leads to results that surpass the performance of both single techniques.**

I. Introduction

Using stereoscopic depth for obstacle detection or road surface estimation has been and still is quite popular [1], [2], [3], [4], [5], [6], [7]. One of the most frequently used depth estimation techniques is block-matching because it is lightweight and easy to implement, which is of great advantage in real-time applications with limited memory and computing power.

A common assumption of all stereo calculation methods is that a physical point in the 3-D world should have a similar appearance, when viewed from different vantage points. Hence, the basic idea is that using two cameras the same 3-D point can be found by correlating left and right stereo image pixels. The similarity is defined by the matching cost function used for the correlation. Typically, single pixel correlations are too error prone so that the correlations of multiple pixels have to be aggregated. In block-matching stereo this is done by comparing pixels within corresponding rectangular image patches in the left and right stereo images. Unfortunately, this imposes a further assumption that the spatial arrangement of pixels will not change between the left and right image, as otherwise the correlation might fail. However, it only holds true for surfaces that are fronto-parallel to the (parallel) cameras. Surfaces that are not fronto-parallel to the cameras get geometrically distorted (differently for each camera).

In other words, this means that the inherent assumption of block-matching (or more generally patch-matching) stereo algorithms is that the depth within a matching window is homogeneous. To a lesser extent, the assumption is also inherent in other stereo algorithms, depending on their spatial integration area and the formulation of their smoothness constraint.

[1]Honda Research Institute Europe GmbH
Carl-Legien-Strasse 30, D-63073 Offenbach/Main, Germany
Email: firstname.lastname@honda-ri.de

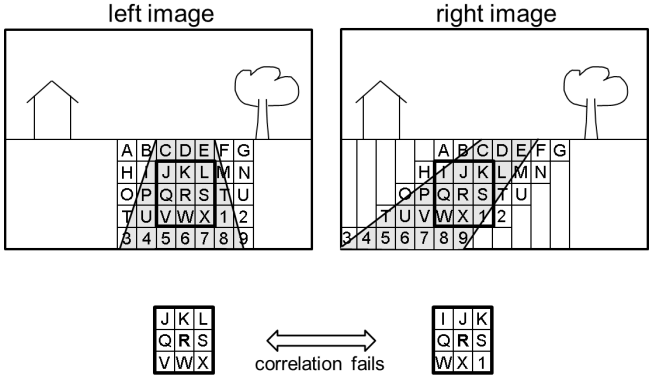


Fig. 1: This figure highlights the problem of the inherent constant disparity assumption of standard patch-matching stereo. In real-world scenarios the disparity values within a patch can vary drastically. One example are highly slanted surfaces as shown here. The non-constant disparity leads to strongly dissimilar corresponding patches. This dissimilarity will lead to a failure of the correspondence search due to a bad correlation value.

In many cases, small deviations from the fronto-parallel assumption do not lead to a severe performance decrease of patch-matching stereo, so that usually this effect is not very prominent, or goes unnoticed within other causes that deteriorate stereo estimation like missing structure contrast or image noise. However, the fronto-parallel assumption is violated very often since a 3-D scene seldomly consists of only upright planar surfaces; and in extreme cases the violation can become too strong to be neglected (see Fig. 1).

As already pointed out in [8], one prominent case is the road surface as it is observed from stereo cameras within a car. Since stereo cameras in a car are typically mounted in a forward-looking position, the optical axes of the stereo rig cameras are (almost) perpendicular to the road's surface normal. Due to this, an image patch of the road will contain a large range of depth values. This will in turn lead to a strong appearance change of the patch between the left and the right stereo camera image which hampers the correspondence search. Fig. 1 explains this illustratively: The left side shows a camera image of a road scene from which a patch is taken and searched for in the right camera image. There, a patch around the same physical point exhibits a very different appearance due to the geometrical distortion generated by the depth gradient, resulting in bad correlation values.

Our idea to counteract the appearance change between the left and right camera image is to relax the fronto-parallel assumption. To this end, we borrow an idea from semi-global matching (SGM) [9]. Instead of just aggregating the pixel-cost for a certain disparity, the best pixel-costs

978-1-4799-3637-3/14/$31.00 ©2014 IEEE

among multiple disparities are computed for each pixel before aggregation. This allows for small depth variations within a patch. In contrast to SGM, the best disparity selection is done only for similar disparities and not across the entire disparity distribution and there is no iterative aggregation along lines. This way we keep the small memory footprint and parallelism of block-matching stereo which is particularly important for low-cost systems.

In the past there have already been some attempts to improve disparity estimation for non-homogeneous surfaces. On the one hand, there are approaches that try to deal with depth discontinuities as encountered at object borders. The most prominent approaches are based on adaptive support weight [10] which uses a weighted aggregation based on the appearance and spatial similarity of the pixels within a patch. The assumption behind this is that pixels belonging to the same surface are close together and share a similar appearance (mainly color). However, these approaches have yet to be shown useful in real-world images which typically are not as nicely colored as test images. Another approach has been introduced in [11]. Here the patches are divided into sub-patches and for matching only the best matching sub-parts are integrated. On the other hand, some approaches try to incorporate surface models [12], [8]. The main disadvantage is that the results easily get worse if the model-assumption (e.g. planar surface) is wrong. Hence, a model-free approach would be more general and suited for a broader range of applications.

II. Relaxed Block-Matching Stereo

As explained in the last section, the depth estimation in real-world scenes can easily be deteriorated because of the geometrical appearance changes between the left and the right stereo image. In this paper, we try to reduce this problem for block-matching stereo by relaxing the fronto-parallel assumption.

The relaxation is realized by considering multiple disparities at the pixel-level and thus multiple possible pixel-level correlations at the same time. Typically patch-matching stereo works as follows. For each disparity, patches from the left image are compared to patches in the right image. The comparison is done by computing a patch-matching cost (or correlation value) between the two patches. In most cases this cost is computed by first computing a pixel-level cost and then aggregating the pixel-level cost over the full patch:

$$C_d = \sum_i c_{i,d} \ , \tag{1}$$

where C_d is the aggregated patch matching cost and $c_{i,d}$ is the pixel-level cost of pixel i in the left patch and its corresponding pixel in the right patch for disparity d (or the other way around). In parallel stereo camera setups corresponding pixels are always on the same image line, thus,

$$c_{i,d} = f_c(p_L(x_i, y_i), p_R(x_i - d, y_i)) \tag{2}$$

where f_c is the pixel-level matching cost function, $p_L(x_i, y_i)$ is pixel i in the left image and $p_R(x_i - d, y_i)$ the corresponding pixel in the right image. The position of the

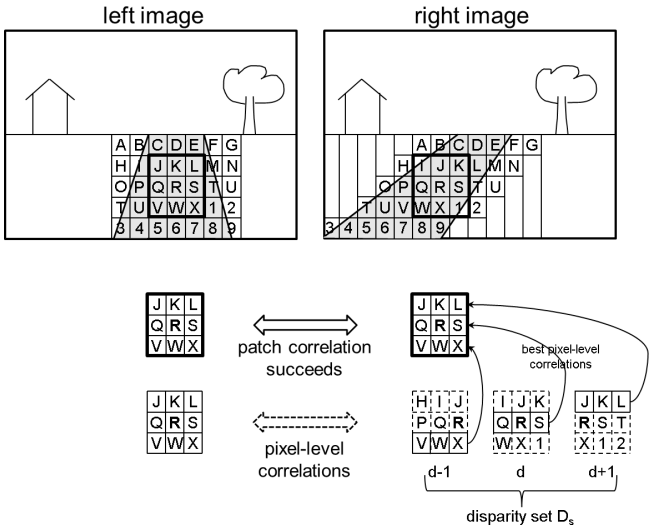


Fig. 2: Relaxed block-matching stereo: Relaxation of the fronto-parallel assumption by taking multiple disparities into account. The top two images are the same as in Fig. 1 where standard block-matching failed to find the correct correspondence. For relaxed block-matching stereo first several pixel-level correlations are computed. Then for each pixel the best pixel-level correlation is detected. In this simple example the lower pixels (VWX) in the patch are best matched with disparity $d-1$, the middle pixels (QRS) are best matched with the patch disparity d and the upper pixels (JKL) are best matched with disparity $d+1$. For the final correlation value only the best correlations per pixel are used. This is depicted here in an illustrative way by creating a new patch.

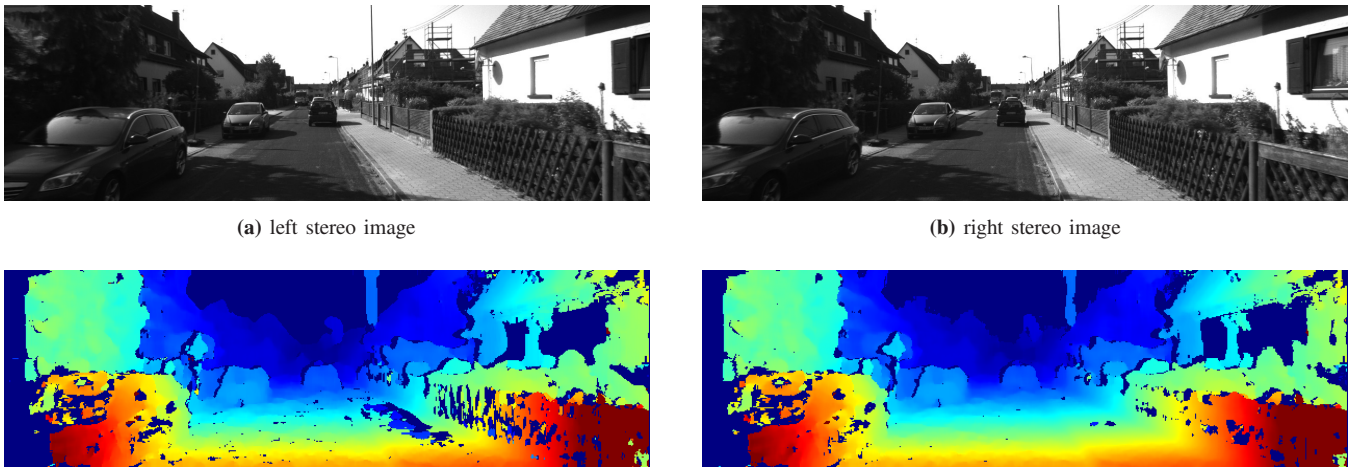
corresponding pixels differs by the disparity d. The pixel-level cost can be a simple *absolute difference* (AD) but it could also incorporate yet another patch (e.g. for computing *normalized cross-correlation* (NCC), *rank transform* (RT) or *census transform* (CT)). The essence of the fronto-parallel assumption in standard stereo processing is that all pixels in a patch have the same disparity d. Hence, for each disparity d the matching-cost C_d of the patch is also computed from the pixel-level matching costs $c_{i,d}$ of disparity d.

In contrast, here we borrow an idea from SGM [9] to not simply aggregate the pixel-level cost of disparity d for the patch disparity d but to incorporate pixel-level costs of multiple disparities by selecting the best pixel-level disparity (the disparity where the pixel-level cost is best) for each pixel from a set of disparities. By this the aggregation equation (1) is extended to

$$C_d = \sum_i \max_{s \in D_s} c_{i,s} \ , \tag{3}$$

where D_s is a set of disparities which can simply be a disparity range like ± 2 disparities, i.e. $D_s = \{d-2, d-1, d, d+1, d+2\}$. The problem of the max selection for finding the best pixel-level disparity is that this might cause plateaus in the patch-level matching cost for neighboring disparities. Hence, it is necessary to penalize pixel-level matching costs that are from a disparity different than the patch disparity,

$$C_d = \sum_i \max_{s \in D_s} P(s,d) c_{i,s} \ , \tag{4}$$



(a) left stereo image

(b) right stereo image

(c) standard block-matching disparity

(d) relaxed block-matching disparity

Fig. 3: This figure shows a side-by-side comparison of standard block-matching stereo and the proposed relaxed block-matching stereo. As can be seen in this example, the relaxed block-matching improves the disparity at the sidewalk and the fence on the right side. Both of these are strongly slanted surfaces.

where P is a penalization that increases with the difference between the patch-level disparity d and the pixel-level disparity s, however, it might also be a fixed penalization for all pixel-level disparities that do not equal the patch-level disparity d.

Figure 2 visualizes the concept in an abstract fashion using the same example as in Fig. 1. While the correlation failed in Fig. 1, now the pixel-level correlation for the patch disparity d and two further disparities, $(d-1)$ and $(d+1)$, are computed first (depicted as the three dashed patches on the lower right). From these pixel-level correlations the best matching for each pixel is selected. In this simplified example the best correlations for the lower pixels are from disparity $(d-1)$, the best for the middle pixels are from the patch disparity d and the best for the upper pixels are from disparity $(d+1)$. These best pixel-level correlations are then combined, possibly with a penalty, to generate a new patch correlation which succeeds. Please note that, although the combination is visualized as building a new patch, the pixel-level correlation values are not used to actually build a new patch. In fact only the pixel-level correlation costs are aggregated to calculate the patch-level correlation cost.

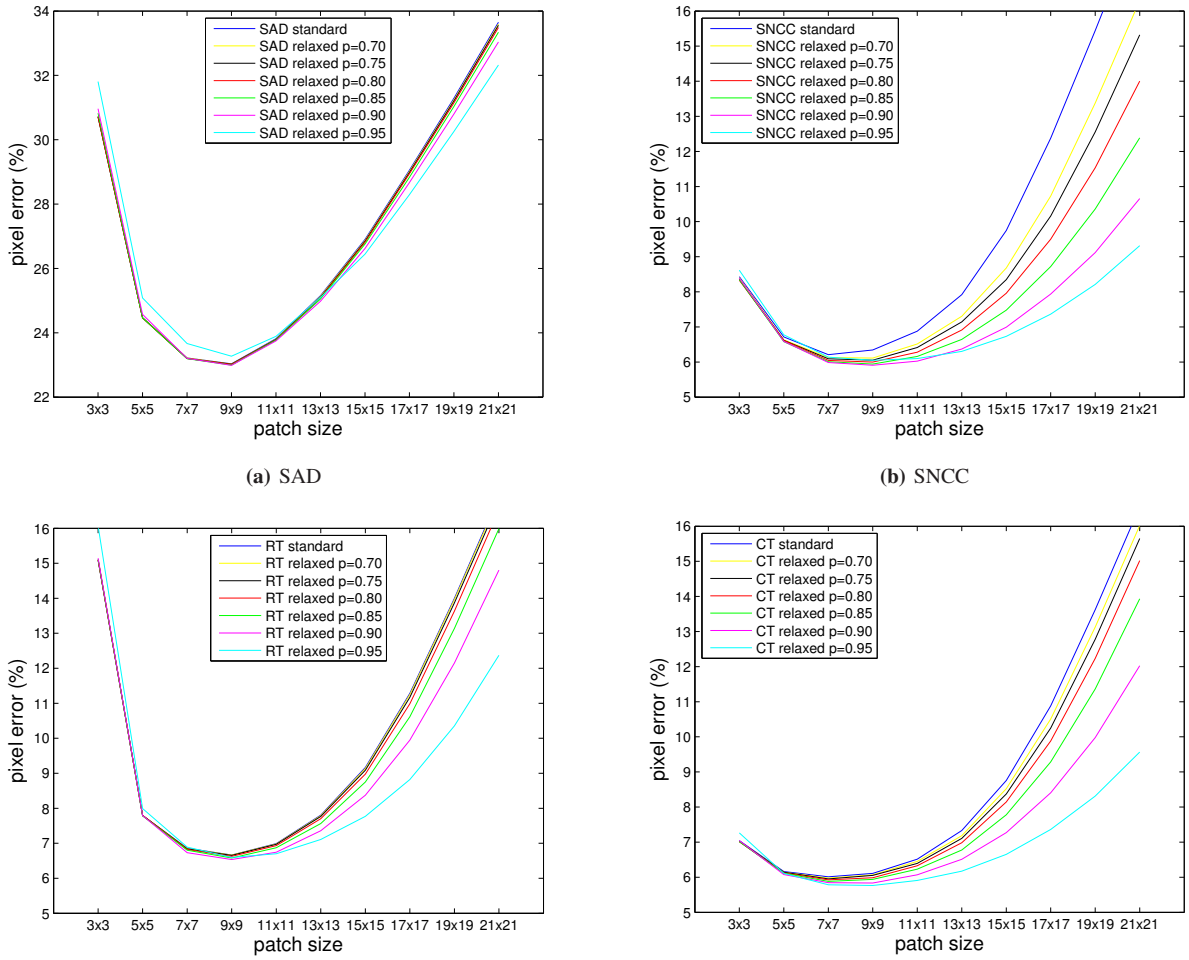
Figure 3 shows the disparity maps of standard block-matching and relaxed block-matching stereo in a side-by-side view. This example demonstrates the improvement of relaxed block-matching stereo. While the standard stereo struggels with some parts of the sidewalk and the fence on the right side, our relaxed stereo approach can handle these strongly slanted surfaces much more robustly.

III. Experimental Results

For analyzing the performance of our approach we used the KITTI benchmark [13] which, among other things, provides stereo images from traffic scenes together with ground truth depth data. As in [8], we found that the best results for block-matching stereo are achieved for half the images size and Gauss ($\sigma = 0.9$) pre-filtered images. For the actual (block-matching) stereo calculation we use an implementation that is similar to the approach of Fua [14], which comprises six steps. Firstly, the patch matching cost C_d is calculated for all pixels and all disparities. In case of the standard block-matching C_d is calculated by the standard aggregation equation (1) and in case of the relaxed block-matching C_d is calculated by the relaxed aggregation equation (4). From these the best matching patch and it's according disparity is selected (winner-takes all). In a second step follows the interpolation of the disparity to sub-pixel accuracy by fitting a quadratic curve to the matching scores in the neighborhood of the optimum. The third step is a left-right consistency check for detecting occlusions and mismatches, i.e. matches that do not pass the left-right check are considered as wrong. Furthermore, in a fourth step small disparity segments are removed. We consider segments with an area smaller than 200 pixels as invalid. In the fifth step, invalidated pixels are filled by interpolating from the next left and next right neighbor that have a valid entry. The final step, applies a 1x9 window and a 9x1 window median filter to iron out small irregularities.

One might argue that comparing stereo algorithms without post-processing would constitute a more direct assessment. However, as was already pointed out in [15], where the same kind of post-processing was used for the assessment of a large set of cost functions, the post-processing reduces the overall error. Without this reduction the large errors impede a fair comparison, while the post-processing increases the discrimination between the different stereo approaches. Furthermore, we use exactly the same post-processing for both the standard and relaxed block-matching stereo which renders the gained results comparable.

In a first evaluation, we investigated how the proposed relaxation improves the stereo disparity estimation compared to the standard block-matching approach. For this test we used the training set of the KITTI stereo benchmark which consists of 194 stereo image pairs and corresponding ground



(a) SAD **(b)** SNCC

(c) RT **(d)** CT

Fig. 4: These results (best viewed in color) show that the relaxation of the fronto-parallel assumption in block-matching for the different cost functions *summed absolute difference* SAD, *summed normalized cross-correlation* SNCC, *rank transform* RT and *census transform* CT. For all four cases the penalty was varied between 0.70 and 0.95. Furthermore, the standard block-matching performance is plotted for comparison. What strikes most is that for SAD there is hardly an improvement while SNCC, RT and CT gain a lot. Additionally, these results show that the relaxed stereo is particularly effective for larger patch sizes. However, an improvement is noticeable for almost the entire size range, including small patch sizes.

truth maps and applied the block-matching stereo with four different patch-level cost functions *summed absolute difference* SAD, *summed normalized cross-correlation* SNCC [16], *rank transform* RT and *census transform* CT [17]. The common property of these cost functions is that they aggregate local measurements (pixel-level measurement) within a larger patch. While in SAD the local measurement is really a pixel-wise measurement, i.e. intensity difference of pixels, the local measurements in SNCC, RT and CT require a certain neighborhood to be computed. In SNCC the pixel-level measurement is a small NCC correlation using a 3x3 neighborhood around each pixel. This is necessary since the calculation of a correlation value requires a variance which cannot be calculated from a single pixel value. For the rank transform the pixel-level measurement itself is pixel-wise (absolute difference of pixel rank); however, for creating the rank per pixel a 11x11 neighborhood is used. Furthermore, census transformed images store a binary fingerprint at each pixel which describes a 7x7 neighborhood around the pixel. Please note that the mentioned neighborhood sizes can vary, we just found these sizes to yield the best results for standard block-matching stereo.

For relaxed stereo we used a disparity set of ± 2 disparities, i.e. $D_s = \{d-2, d-1, d, d+1, d+2\}$. The penalty is 0 for d, p for $d-1$ and $d+1$ and p^2 for $d-2$ and $d+2$. Note that we did not do an extensive test of different kinds of penalties; however, in some pre-tests it turned out that for larger disparity sets there is hardly an improvement but rather a deterioration. Here we concentrate more on the general evaluation of the relaxed stereo and postpone the optimization of the penalty to future work.

Figure 4 shows the result for the four tested cost-functions and their relaxed extension for penalties p from 0.7 to 0.95. Please note that for $p = 0$ the relaxed block-matching becomes the standard block-matching stereo and for $p = 1.0$ there is no penalization at all. Furthermore, all cost-functions

have been reformulated in a way that the cost has to be maximized.

What strikes first is that for SAD there is hardly an improvement while SNCC, RT and CT are strongly improved. The reason might be that SAD is less sensitive to actual structure information because it incorporates no intensity normalization. Due to this the cost difference between neighboring disparities changes much more smoothly as compared to the other three costs which renders the relaxation idea less effective for SAD. However, notice that SAD has an overall much poorer performance than the SNCC, RT and CT variants. The second major observation is that SNCC, RT and CT show a very similar improvement for the same penalty. In all three cases the performance improves until $p = 0.9$. For $p = 0.95$ the results get worse than the standard block-matching for small patch sizes. Altogether it is also obvious that the improvement is best for the larger patch sizes but there is also a significant improvement even for smaller patch sizes. For example standard SNCC with patch size 9x9 improves from a disparity error 6.34% to 5.91% for relaxed block-matching stereo with penalty $p = 0.90$.

In a second evaluation, we compare *relaxed block-matching* RBM stereo against *appearance-aligned block-matching* AABM [8]. The AABM approach also tackles the fronto-parallel assumption. Instead of relaxing the correlation, AABM incorporates some scene knowledge to pre-warp the stereo images in order to compensate for large perspective effects. For example for vehicle applications knowledge about the relative orientation of camera and ground surface can be used to compensate for the large perspective effects of the ground surface, e.g. the street. Figure 5 shows standard block-matching, AABM and relaxed block-matching in comparison. For the comparison we used SNCC as cost function because it was also used in [8] and thus allows for a more meaningful analysis. This comparison demonstrates that relaxed block-matching is nearly as effective as AABM although no model knowledge is used. This is very important because there are many situation were no known models can be applied to the scene in order to compensate for the perspective effects. Hence, the relaxed block-matching proves to be a more general approach. It has to be mentioned though that AABM is much better than RBM for large patch sizes. The reason might be that here we used just ± 2 disparities for relaxing the correlation. It might be necessary to use a larger disparity set for larger patch sizes. This has to be investigated in future work. On the other hand, a combination of both approaches leads to even better results (cyan line with diamonds). For this combination we just straightforwardly applied the relaxed correlation to the pre-warped images used for AABM. This also shows how easy both approaches can be integrated into already existing algorithms.

In a final evaluation, we used the testing data set of the KITTI benchmark to upload our results to the online table (`http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php`) for comparison with other methods. Currently (December 2013), our method

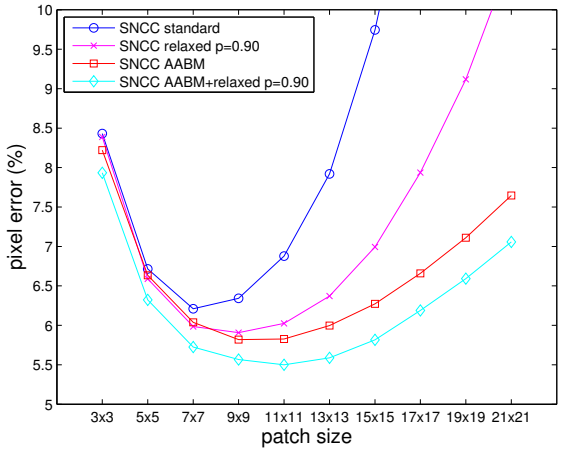


Fig. 5: This plot compares the performance of *standard block-matching* stereo, *relaxed block-matching* RBM and *appearance-aligned block-matching* AABM [8]. We used the same parameters as in [8] for AABM. The comparison highlights that RBM is nearly as effective as AABM although it does not use model knowledge. Furthermore, the results demonstrate that straightforward combination of AABM and RBM lead to even better results than the single approaches.

TABLE I: Snapshot of the online KITTI stereo benchmark table from December 2013. It shows the top three ranking methods at that time: *SceneFlow*, *Particle Convex Belief Propagation with StereoSLIC* PCBP-SS [18], *Grouping-Transfer RF* gtRF-SS; plus some selected methods: *Iterative Semi-Global Matching* iSGM [19], *Total Generalized Variation without Warping* ALTGV [20], *OCV-SGBM-grad census* OCV-SGBM2 [21]; together with four variants of block-matching using SNCC cost functions: *standard block-matching* SNCC [16], *Relaxed Block-Matching* RBM, *Appearance-Aligned Block-Matching* AABM [8] and the combined *Appearance-Aligned Relaxed Block-Matching* AARBM. Note that the block-matching methods, although a bit worse in accuracy, are more than 2 magnitudes faster than the top ranked methods.

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All	Time
1	SceneFlow	2.98%	3.86%	0.8 px	1.0 px	5min
2	PCBP-SS	3.40%	4.72%	0.8 px	1.0 px	6min
3	gtRF-SS	3.83%	4.59%	0.9 px	1.0 px	1min
...						
10	iSGM	5.11%	7.15%	1.2 px	2.1 px	8s
11	**AARBM**	**5.14%**	**6.20%**	**1.1 px**	**1.2 px**	**0.40s**
12	ALTGV	5.36%	6.49%	1.1 px	1.2 px	20s
13	OCV-SGMBM2	5.38%	6.50%	1.0 px	1.2 px	2s
14	AABM	5.42%	6.52%	1.1 px	1.3 px	0.43s
15	**RBM**	**5.50%**	**6.48%**	**1.2 px**	**1.3 px**	**0.20s**
...						
21	SNCC	6.22%	7.28%	1.4 px	1.5 px	0.27s

relaxed block-matching RBM ranks 15th among 46 algorithms in the online KITTI stereo benchmark. Moreover, the combined approach of *apperance-aligned relaxed block-matching* ranks 11th among the 46 algorithms. Table I displays a selection of methods in the KITTI benchmark together with standard block-matching SNCC, relaxed block-matching RBM, appearance aligned block-matching AABM and appearance aligned relaxed block-matching AARBM from the KITTI benchmark in December 2013.

Note that the strength of our approach is the good trade-off between speed and accuracy. This is also reflected in Fig. 6 which shows the performance of all algorithms in the KITTI benchmark in December 2013 together with their runtime.

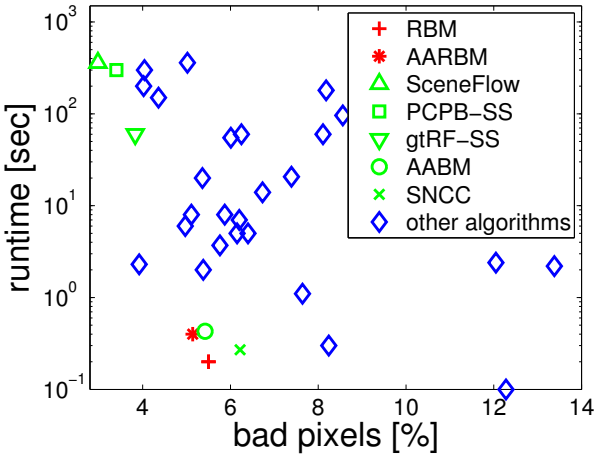


Fig. 6: This plot compares the performance of all algorithms in the KITTI benchmark in December 2013 against their runtime. This shows that our relaxed block-matching scheme (RBM and AARBM) features a good trade-off between speed and accuracy. Please note the logarithmic scale of the runtime.

IV. Conclusion

In this paper, we presented a variant of *block-matching* BM stereo that relaxed the inherent fronto-parallel assumption of BM in order to improve the correlations between the two stereo images. The underlying idea is to change the aggregation procedure of BM stereo. Instead of summing pixel-level costs of only one disparity for computing patch correlations, first the best pixel-level costs of a set of disparities are computed for each pixel. This way the disparities within one patch do not need to be equal in order to yield a good correlation.

We conducted several experiments using the KITTI benchmark. First, we demonstrated that for the cost functions *summed normalized cross-correlation* SNCC, *rank transform* RT and *census transform* CT that there is a significant improvement using our relaxation scheme. However, the experiment also showed that there is only a minor improvement for *summed absolute difference* SAD. Second, we compared the relaxed scheme with *appearance-aligned block-matching* AABM which also tackles the problem of the inherent fronto-parallel assumption in BM stereo. We could show that for smaller patch-sizes both approaches yield similar results, however that AABM is better for larger patch sizes. On the other hand, we could show that a combination of both approaches leads to results that outperform both techniques in isolation. For a more transparent comparison we also uploaded our results into the online KITTI benchmark. As of December 2013 we ranked 15th and 11th with relaxed block-matching and relaxed appearance-aligned block-matching, respectively, among 46 approaches. However, in contrast to higher ranked approaches our algorithm runs at least five times faster which highlights the good trade-off between speed and performance which is especially important for low-cost systems in autonomous vehicles.

References

[1] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 62–81, 1998.

[2] A. Seki and M. Okutomi, "Robust obstacle detection in general road environment based on road extraction and pose estimation," in *Proceedings of the Intelligent Vehicles Symposium (IV)*, 2006, pp. 437–444.

[3] F. Oniga, S. Nedevschi, M. M. Meinecke, and T. B. To, "Road surface and obstacle detection based on elevation maps from dense stereo," in *Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2007, pp. 859–865.

[4] P. Steingrube, S. K. Gehrig, and U. Franke, "Performance evaluation of stereo algorithms for automotive applications," in *Proceedings of the 7th International Conference on Computer Vision Systems: Computer Vision Systems (ICVS)*, 2009, pp. 285–294.

[5] J. Schmüdderich, N. Einecke, S. Hasler, A. Gepperth, B. Bolder, R. Kastner, M. Franzius, S. Rebhan, B. Dittes, H. Wersing, J. Eggert, J. Fritsch, and C. Goerick, "System approach for multi-purpose representations of traffic scene elements," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2010, pp. 1677–1684.

[6] M. Bertozzi, L. Bombini, A. Broggi, M. Buzzoni, E. Cardarelli, S. Cattani, P. Cerri, A. Coati, S. Debattisti, A. Falzoni, R. I. Fedriga, M. Felisa, L. Gatti, A. Giacomazzo, P. Grisleri, M. C. Laghi, L. Mazzei, P. Medici, M. Panciroli, P. P. Porta, P. Zani, and P. Versari, "VIAC: An out of ordinary experiment," in *Proceedings of the Intelligent Vehicles Symposium (IV)*, 2011, pp. 175–180.

[7] C. Unger, E. Wahl, and S. Ilic, "Efficient stereo and optical flow with robust similarity measures," in *33rd Annual Symposium of the German Association for Pattern Recognition*, 2011, pp. 246–255.

[8] N. Einecke and J. Eggert, "Stereo image warping for improved depth estimation of road surfaces," in *Proceedings of the Intelligent Vehicles Symposium (IV)*, 2013, pp. 189–194.

[9] H. Hirschmüller, "Stereo processing by semi-global matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2008.

[10] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 650–656, 2006.

[11] P. R. I. Heiko Hirschmller and J. M. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," *International Journal of Computer Vision (IJCV)*, vol. 47, no. 1-3, pp. 229–246, 2002.

[12] M. Habbecke and L. Kobbelt, "Iterative multi-view plane fitting," in *Proceedings of the 11th International Fall Workshop Vision, Modeling, Visualization (VMV)*, 2005, pp. 73–80.

[13] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3354–3361.

[14] P. Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," *Machine Vision and Applications*, vol. 6, no. 1, pp. 35–49, 1993.

[15] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1582–1599, September 2009.

[16] N. Einecke and J. Eggert, "A two-stage correlation method for stereoscopic depth estimation," in *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2010, pp. 227–234.

[17] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proceedings of Third European Conference on Computer Vision (ECCV)*, 1994, pp. 151–158.

[18] K. Yamaguchi, D. McAllester, and R. Urtasun, "Robust monocular epipolar flow estimation," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 1862–1869, 2013.

[19] S. Hermann and R. Klette, "Iterative semi-global matching for robust driver assistance systems," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2012.

[20] G. Kuschk and D. Cremers, "Fast and accurate large-scale stereo reconstruction using variational methods," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, 2013, pp. 1–8.

[21] "Anonymous KITTI submission."