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

Stereo Image Warping for Improved Depth Estimation of Road Surfaces

Nils Einecke, Julian Eggert

2013

Preprint:

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

2013 IEEE Intelligent Vehicles Symposium (IV)
June 23-26, 2013, Gold Coast, Australia

Stereo Image Warping for Improved Depth Estimation of Road Surfaces

Nils Einecke and Julian Eggert
Honda Research Institute Europe GmbH
63073 Offenbach/Main, Germany
{nils.einecke,julian.eggert}@honda-ri.de

Abstract— **Accurate stereoscopic depth estimation, in particular of the road surface area, is one of several key technologies to improve Advanced Driver Assistance Systems (ADAS). One major problem is that the quality of the stereoscopic depth measurements of the road is often poor - which is mainly attributed to a lack of texture on the road surface. Especially for patch-matching stereo algorithms, the estimated depths look irregular and bumpy. In this paper, we show that the violation of the fronto-parallel assumption is the major reason for a bad depth estimation and not a low-contrast texture on the road surface. Since patch-matching or block-matching stereo inherently assumes a constant disparity within one patch, this is violated if the cameras are oriented almost parallel to the ground, which is typically the case in ADAS, and which leads to a strong distortion of the appearance between the two cameras. In order to tackle this problem, we propose a compensation of this distortion by applying a linear warp on one of the stereo images according to the expected disparity for the planar ground. This recovers the fronto-parallel assumption and results in a very good depth estimations of road surfaces. Our experiments on the KITTI stereo benchmark demonstrate the quantitative competitiveness of the approach, while retaining the speed and simplicity of block-matching stereo approaches. Furthermore, our experiments show that the approach is very robust, achieving results for the road surface that are significantly better than standard patch-matching stereo processing without warping for a wide range of warp parameter settings.**

I. INTRODUCTION

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

Common to all stereo calculation methods is the assumption that the appearance around the same physical point, if looked at by two different cameras, should have a similar brightness structure (appearance). In particular in calibrated stereo arrangements with parallel cameras, this is true on surfaces that are fronto-parallel to the cameras. However, for surfaces that are not fronto-parallel to the cameras, the projected images get geometrically distorted (differently for each camera).

Stereo algorithms based on patch-matching are particularly susceptible to these distortions since they compare the appearances around a physical point as a whole. As long as the surface from which the structure has been captured lies fronto-parallel to both cameras all pixels within a patch (or block, if of rectangular shape) in one camera match exactly the brightness structure (appearance) of a corresponding patch at the second camera. This means that the inherent assumption of patch-matching algorithms is that the physical structure behind the area of the matching patches should have the same depth in order to find good correspondences. To a lesser extent, the argument applies to all stereo matching algorithms, depending on their spatial integration area and the formulation of their smoothness constraint.

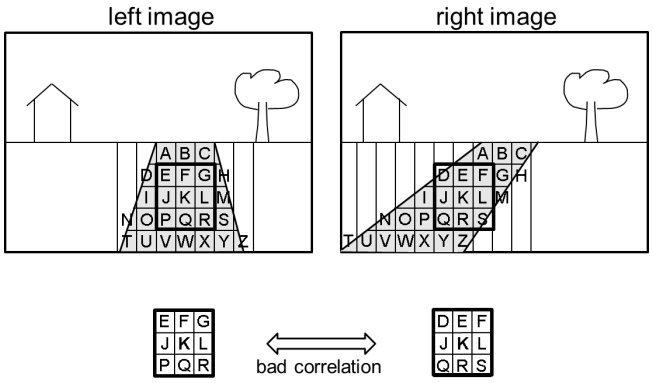


Fig. 1: This figure highlights the problem of the inherent constant disparity assumption of standard patch-matching stereo. For highly slanted surfaces, e.g. a street, the disparity values within a patch can vary drastically which leads to strongly dissimilar corresponding patches. This dissimilarity will lead to a failure of the correspondence search due to a bad correlation value.

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, as it is the case for missing structure contrast or large intensity noise levels. However, the fronto-parallel assumption is violated very often since a 3-D scene seldomly consists of only upright planar surfaces. In extreme cases the violation can become too strong to be neglected (see Fig. 1).

One such 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's 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 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 apply an image warping to one of the stereo images. By doing so the left and right image patches become more similar and allow for a more robust correspondence search. Image warping in stereo processing has seldomly been used in the recent time. However, when the first vision systems for the automotive domain have been proposed, image warping was used to reduce the computational effort of feature-based stereo processing [8], [1]. Using a known geometry of the road surface one stereo image is warped into the other camera's view (or birds eye view). By doing so, obstacles can be detected very fast since road feature points have a near zero disparity in the warped images while non-road features have a residual disparity. Another way of finding road and non-road is to subtract the warped image of one camera from the non-warped image of the other camera [9]. While road pixels will be canceled out, pixels from upright obstacles retain high values. However, in these cases image warping was not used to improve the accuracy of the stereoscopic depth estimation. In contrast, [10], [11] integrate image warping directly into the correspondence search for very accurate depth estimations. The main drawback of this approach is the high computational cost as compared to standard block-matching because the image warping has to be done multiple times for each image patch.

II. APPEARANCE ALIGNED STEREO

As explained in the last section, the depth estimation of a road surface is usually deteriorated, mainly because of the geometrical appearance changes between the left and the right stereo image. In this paper, we introduce a way to correct for these geometrical distortions by means of image warping and analyze the impact on stereo calculation results.

In a parallel stereo camera setup the geometrical appearance change of a planar surface, e.g. a road, between the left and right stereo camera can be described by means of an affine transformation. If we apply the inverse of this transformation to one of the stereo images, then the appearance of the road in the left and the right stereo image becomes more similar. In this paper we apply the image transformation ω to the right stereo image R to get the warped image R^T

$$R^T = \Omega(R) \ . \tag{1}$$

This means that the position (x, y) of a pixel p is changed to (x^T, y^T) via the position transformation ω

$$x^T = \omega_x(x, y) \tag{2}$$
$$y^T = \omega_y(x, y) \ . \tag{3}$$

Note that in a parallel stereo setup $\omega_y(x, y) = y$ since there can only be distortions along the x-direction.

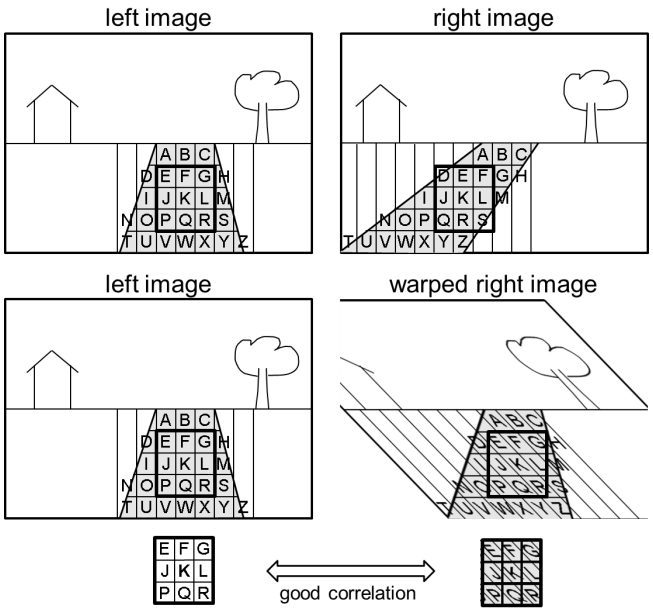


Fig. 2: Block-matching stereo with appearance alignment: Compensation of the geometrical distortion of stereo images for slanted surfaces. Top row: Original left and right stereo camera images. Bottom row: The warping of the right image compensates the geometrical patch distortion, so that a patch correspondence search succeeds here.

The expected result is that the correspondence search between these appearance aligned stereo images is improved on a per-patch basis, meaning that corresponding patches should now exhibit better matching values. This effect is schematically shown in Fig. 2. On the top row, we see the original left and right camera images with the geometrical distortion on the road plane. On the bottom row, we see the original left image together with the warped right image. The distortion of the patch region now has been compensated, and a patch correspondence search succeeds.

As was mentioned above, we apply the *appearance alignment* in form of a geometrical transformation to the right stereo image and then calculate the depth with respect to the left stereo image. Since the appearance alignment changes the x-position of the pixels in the right image this also introduces a systematic distortion in the disparity results according to the same transformation. Consider a correspondence pixel pair p_l and p_r in the left and right image, respectively. The position of p_L is (x_L, y_L) and the position of p_R is (x_R, y_R), where

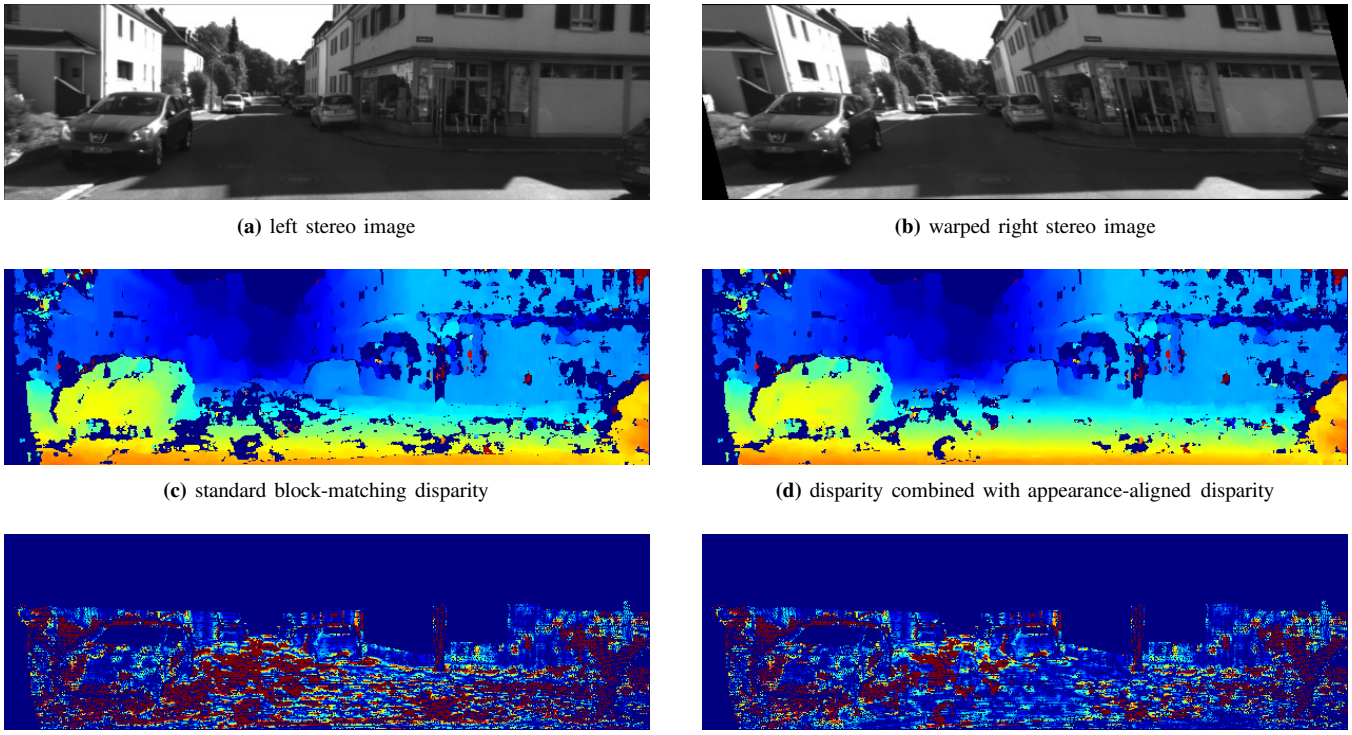
$$x_R = x_L - d(x_L, y_L) \tag{4}$$

i.e. the position of p_R is shifted by the disparity $d(x, y)$ in the right image. Due to the geometrical transformation ω_x the position of the right pixel is changed to

$$x_R^T = \omega_x(x_R, y_R) \tag{5}$$

This leads to a decrease of the disparity $d_a(x, y)$ estimated with appearance alignment according to the x-shift s_x introduced by ω_x

$$d_a(x_L, y_L) = d(x_L, y_L) - s_x(x_R, y_R) \ , \tag{6}$$



(a) left stereo image

(b) warped right stereo image

(c) standard block-matching disparity

(d) disparity combined with appearance-aligned disparity

(e) standard block-matching disparity error

(f) disparity combined with appearance-aligned disparity error

Fig. 3: Comparison of standard block-matching stereo (left column) and the combination of standard and appearance-aligned stereo (right column). The top images show the left stereo image and the aligned (warped) right stereo image. In the middle row the images show disparity maps with disparity coded in colors from blue (far) to red (close). The bottom images show disparity error maps where red means large error and blue means low error. For this example frame 19 from the KITTI test benchmark [12] was used. It demonstrates the improvement in density as well as in accuracy of the estimated disparities for combining the standard stereo approach with the appearance-aligned stereo. While the standard disparity map shows some very large holes on the street and gross disparity errors, the combined disparity map of normal and appearance-aligned stereo is more dense and much smoother on the street surface. Since the combination uses the best disparity (highest correlation) from each disparity map the disparity for scene parts other than the road are mainly from the normal disparity map. This way only the positive effects of the image warping take effect.

where the x-shift s_x is

$$s_x(x,y) = x_R^T - x_R = x_R^T - \omega_x^{-1}(x_R^T, y_R^T) \ . \qquad (7)$$

Therefore, it is important to correct the resulting disparity map d_a for the transformation applied to the right stereo image. It is obvious from the shift equation (6) that this correction reduces to a simple addition.

$$\begin{aligned} d(x_L, y_L) &= d_a(x_L, y_L) + s_x(x_R, y_R) \qquad (8) \\ &= d_a(x_L, y_L) + x_R^T - \omega_x^{-1}(x_R^T, y_R^T) \ . \qquad (9) \end{aligned}$$

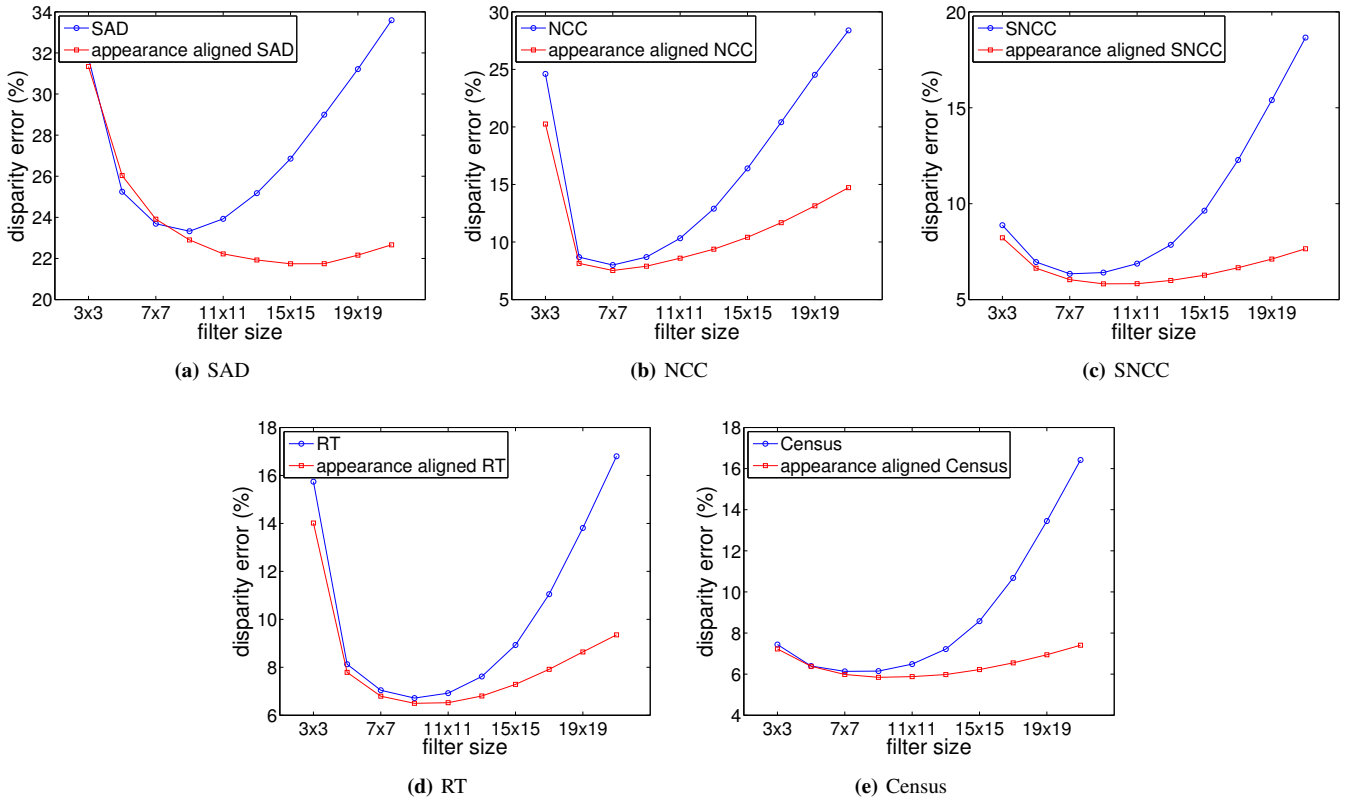
One drawback of this approach so far is that usually it is not known in which regions of the image the warping model has to be applied since the layout of the scene (resp. the orientation of the surfaces) is usually not known in advance. The naïve way for scenes with a predominant surface orientation is to apply the warping to the full image and hope that everything works out. However, warping the full image can violate the fronto-parallel assumption for non-road scene parts which will decrease the stereo performance for those scene parts.

To solve this problem, in an extension of the method we propose to compute two separate disparity maps, one for the unmodified stereo images and one for the appearance aligned stereo images, and then combine the disparity map results on a pixel-wise basis depending on a confidence measure. That is, for each pixel, from the two disparity maps we choose e.g. the disparity of the map with the highest correlation coefficient.

Altogether, our approach can be summarized as a sequence of six simple steps:

1) stereo image acquisition
2) disparity computation using the normal images
3) appearance alignment by warping the right camera image
4) disparity computation using the appearance aligned images
5) correction of the appearance aligned disparity map
6) combination of the normal disparity map and the corrected appearance aligned disparity map

Fig. 3 shows the disparity map and disparity error map for block-matching stereo with and without appearance alignment. It demonstrates that the combination with appearance-aligned disparity improves disparity maps in density as well as in accuracy. While the standard disparity map shows some very large holes on the road and gross disparity errors, the combined disparity map with the appearance-aligned stereo is more dense and much smoother on the road surface. For scene parts other than the road both approaches yield



(a) SAD **(b)** NCC **(c)** SNCC

(d) RT **(e)** Census

Fig. 4: Comparison of standard block-matching stereo to the proposed appearance-aligned stereo based on image warping for the five different cost functions: *summed absolute difference* SAD, *normalized cross-correlation* NCC, *summed normalized cross-correlation* SNCC, *rank transform* RT and Census transform. For this comparison the test set of the KITTI benchmark [12] was used. The plots show that for all analyzed cost functions the appearance alignment leads to a significant improvement of the disparity error. Please note that disparity error is computed for the full images and not just the road surface.

almost the same disparity values which demonstrates that the appearance-alignment stereo has no negative influence on scene parts with a different transformation.

Please note that in the remainder of the paper the term *appearance-aligned stereo* will always refer to all six steps, i.e. it refers always to the combination of the normal disparity map with the appearance-aligned disparity map.

III. EXPERIMENTAL RESULTS

For analyzing the performance of our approach we used the newly implemented KITTI benchmark [12] which, among other things, provides stereo images from traffic scenes together with ground truth depth data. For the warping model we use a very simple linear warping that shifts each image line according to its vertical image position

$$x^T = \omega_x(x,y) = x + ay + n \quad (10)$$

$$y^T = \omega_y(x,y) = y\ . \quad (11)$$

In this equation $\omega_x(x,y)$ is the shift applied to the image line y. The parameters a and n are the gain and the bias of this linear shift model. In particular a indirectly describes the pitch angle between the ground surface and the stereo camera.

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 [13], which comprises six steps. Firstly, the matching cost is calculated for all pixels and all disparities. From these the best matching 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 5x5 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 [14], 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 normal and appearance aligned

stereo which renders the gained results comparable.

In a first experiment, we investigated how the appearance alignment improves the stereo disparity estimation of 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 truth maps. Fig. 4 shows the performance of block-matching stereo for the five different cost functions *summed absolute difference* SAD, *normalized cross-correlation* NCC, *summed normalized cross-correlation* SNCC [15], *rank transform* RT and Census transform [16]. For each cost function ten different filter sizes ranging from 3x3 to 21x21 were tested. SNCC, RT and Census are two-stage methods which in a first stage calculate a small-windowed correlation (3x3 window), windowed rank transform (11x11 window) and windowed Census transform (7x7 window) and in a second stage integrate the results; here the filter sizes correspond to the integration stage. In this experiment the parameters of the linear model were extracted from the first disparity map, i.e. which was computed from the first stereo image pair, and then fixed ($a = 0.26$ and $n = -23$) for the whole data set.

In each plot of Fig. 4, the performance of the standard block-matching stereo (blue line with circles) is compared to the appearance-aligned stereo using image warping (red line with squares). The plots demonstrate that for all cost functions the appearance-aligned stereo leads to a significant improvement of the disparity error. It also strikes that the optimal filter sizes are shifted to larger sizes for all cost functions except NCC. This confirms that the appearance alignment indeed increases the similarity of corresponding patches so that larger patches with larger discrimination power can be used reliably without breaking the assumption of homogeneous disparity.

One question that remains is how critical the choice of the parameters of the warp model are. In order to answer this question we scanned the parameter space of our simple linear model using SNCC block-matching stereo. As explained above, in this model the gain a describes the change of the disparity in vertical image direction, i.e. a somehow describes the pitch between the stereo camera and the planar ground surface. Since the bias n will only change the search range of the stereo algorithm it is not investigated further. Fig. 5 plots the performance of the appearance-aligned stereo for different choices of the gain a for block-matching stereo with SNCC. It can be seen that the disparity error decreases until an optimal gain of $\sim$0.25 has been reached and then increases again. The curve demonstrates that the choice of the gain is not very sensitive since the error curve is very smooth. It strikes, however, that at some point the error can pass the baseline performance of the standard block-matching stereo without appearance alignment. This means one should be rather conservative when choosing a parameter since an underestimation can at worst lead to the same performance as without appearance alignment.

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 2012), our method (*Appearance-Aligned Block-Matching* AABM) ranks 4th in the online KITTI stereo benchmark. Table I displays the best six methods from the KITTI benchmark in December 2012. Our algorithm has 0.8 percentage points less erroneous pixels (in non-occluded areas) than the best block-matching method in the benchmark which is SNCC. The top three approaches have 0.3 to 1.4 percentage points less erroneous pixels (in non-occluded areas) than our approach. However, the overall average disparity error of our method is very similar or even better than the best ranked methods. In addition, our method is by large factors faster than the top-ranked methods on standard hardware, and close to real-world applicability. It takes 430ms on one core of a 3GHz processor. Hence, appearance-aligned stereo provides a very good trade-off between speed and accuracy which is also supported by the comparative plot in Fig. 6. In the worst case, the proposed method is two times slower than standard block-matching since block-matching has to be called twice, once for the normal stereo

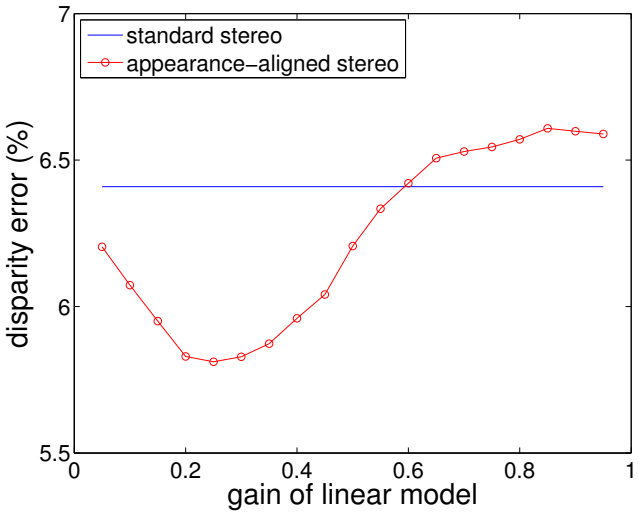


Fig. 5: Performance of the appearance-aligned stereo approach for different parameters of the linear warping model. The plot demonstrates that for a large parameter range the alignment leads to a significant improvement over the standard stereo approach. However, one should not apply to strong warpings as these could lead to slightly worse results as the standard approach. Fortunately, the maximal degradation seems to be bound as the error curve saturates at 6.6%.

TABLE I: Snapshot of the online KITTI stereo benchmark table from December 2012. It shows the top six ranking methods at that time: *Particle Convex Belief Propagation* PCBP [17], *Iterative Semi-Global Matching* iSGM [18] , *Calibrated Total Generalized Variation* CTGV (anonymous submission), our *Appearance-Aligned Block-Matching* AABM, *Semi-Global Matching* SGM [19] and *Summed Normalized Cross-Correlation* SNCC [15]. Note that the proposed method is factors >600, >18 and >800 faster than the first three methods while retrieving comparable disparity errors.

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All	Time
1	PCBP	4.13%	5.45%	0.9 px	1.2 px	5min
2	iSGM	5.16%	7.19%	1.2 px	2.1 px	8s
3	CTGV	5.17%	7.02%	1.1 px	1.7 px	6min
4	**AABM**	**5.50%**	**6.60%**	**1.1 px**	**1.3 px**	**0.43s**
5	SGM	5.83%	7.08%	1.2 px	1.3 px	3.7s
6	SNCC	6.27%	7.33%	1.4 px	1.5 px	0.27s

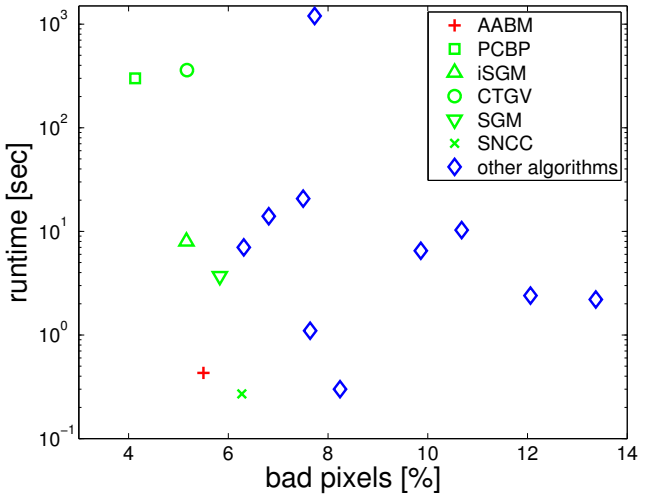


Fig. 6: Runtime vs. accuracy for the 17 best algorithms of the KITTI stereo benchmark table (December 2012). This comparison highlights the good trade-off between speed and accuracy that appearance-aligned block-matching stereo (AABM) provides. For the abbreviations please see Table I.

images and once for the appearance aligned images. This could be improved further because the appearance alignment also reduces the search space.

IV. SUMMARY

In this paper, we presented an extension to the standard patch-matching stereo algorithms for improving the depth estimation of the ground surface. We argue that the problem of bad stereo matches on ground surfaces is mainly caused by the strong violation of the patch-matching inherent fronto-parallel assumption. Since the ground plane's normal vector is typically perpendicular to the stereo cameras optical axis the appearance of patches in the left and right camera change drastically which in turn hampers the correlation process.

In our proposed appearance-aligned stereo we reduce this problem by applying image warping to the stereo camera images in order to make them more similar in appearance. The conducted experiments demonstrate that for all tested cost function the appearance alignment leads to a significant improvement with little extra computational cost. Furthermore, our experiments show that the choice of the model parameters is very stable so that there is no need for highly accurate model fitting. Indeed we could show a significant improvement of block-matching stereo with a fixed parameter setting for the appearance alignment on the KITTI benchmark which consists of street scenario images of different complexity and structure.

In future work we will investigate how the appearance-aligned stereo can be improved further. One idea is to reduce the search space for the appearance aligned images since the alignment leads to a reduction of disparity differences. Another target is to apply the appearance alignment idea to patch-based optical flow. In contrast to the stereo case, the parameters of the warping model cannot be fixed for the optical flow case since they depend on the movement of the camera between two time steps, i.e. they depend on the speed of the vehicle.

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 2006 IEEE 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 *Proceedings of the 33rd Annual Symposium of the German Association for Pattern Recognition*, 2011, pp. 246–255.

[8] Q.-T. Luong, J. Weber, D. Koller, and J. Malik, "An integrated stereo-based approach to automatic vehicle guidance," in *Proceedings of the Fifth International Conference on Computer Vision*, 1995, pp. 52–57.

[9] M. Okutomi, K. Nakano, J. Maruyama, and T. Hara, "Robust estimation of planar regions for visual navigation using sequential stereo images," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2002, pp. 3321–3327.

[10] 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.

[11] N. Einecke, S. Rebhan, V. Willert, and J. Eggert, "Direct surface fitting," in *Proceedings of the Fifth International Conference on Computer Vision Theory and Applications (VISAPP)*, 2010, pp. 125–133.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] K. Yamaguchi, T. Hazan, D. McAllester, and R. Urtasun, "Continuous markov random fields for robust stereo estimation," in *Proceedings of Third European Conference on Computer Vision (ECCV)*, 2012, pp. 45–58.

[18] 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.

[19] 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.