Illumination Invariant Dense Image Matching based on Sparse Features

MAX MEHLTRETTER[1] & CHRISTIAN HEIPKE[1]

Abstract: In this paper we propose an algorithm for dense depth estimation based on a cost function, which is robust against changes in illumination. For this purpose, we utilize census transformation and Phase Congruency. The process of cost computation is supported by a triangle-based depth prediction approach using a set of matched feature points. These points can reliably be detected even under drastic changes in illumination. We demonstrate the performance of our approach on the challenging KITTI stereo dataset and a set of images with simulated changes. The results show that the concept achieves state-of-the-art on images with similar illumination conditions and is robust against changed conditions.

1 Introduction

Reliable and up-to-date maps received much attention in the field of mobility in recent times (SEIF et al. 2016). Especially the sector of assisted and autonomous driving strongly depends on such maps in order to be able to accurately determine the vehicle's pose and compute a safe trajectory. With regard to the importance of the currency of the information contained in such a map, data acquisition must be done continuously.

For the purpose of data acquisition, mobile mapping using stereo cameras is well suited. In comparison to laser scanners, stereo cameras are less expensive and furthermore, they do not only provide geometric, but also multi-spectral radiometric information. Hence, additional information for later image segmentation and classification is provided. But, special attention has to be paid to the fusion of current information and data captured earlier to obtain a consistent, reliable and accurate result. Depending on the temporal distance between two recordings, changes in illumination can lead to significant differences in the acquired data. Not only the color of illumination or the geometric relationship between illumination, scene and sensor but also the number of light sources might have changed, making the fusion task a potentially difficult challenge.

In this paper, we present a novel approach for illumination invariant dense image matching. Following the quote of Aristotele “The whole is greater than the sum of its parts”, we propose a cost function combining different approaches to achieve robustness against changes in illumination. For this purpose, we merge cost functions based on census transformation (ZABIH et al. 1994) and based on images transformed with Phase Congruency (KOVESI 1999). The disparity computation is supported by a triangle-based depth prediction approach (BULATOV et al. 2011) utilizing a set of matched feature points which are reliably detectable even under drastic changes in illumination (VERDIE et al. 2015). The optimal solution for the disparity image is determined by semi-global matching (HIRSCHMÜLLER 2008).

[1] Leibniz Universität Hannover, Institute of Photogrammetry and GeoInformation, Nienburger Straße 1, D-30167 Hannover, E-Mail: [mehltretter, heipke]@ipi.uni-hannover.de

The remaining sections of this paper are organized as follows: In section 2, we provide an overview of the field of illumination invariant image matching. Section 3 contains the methodology of our approach. Then, in Section 4, the experimental results are shown and explained. Finally, the conclusion is presented in Section 5.

2 Related Work

The reconstruction of depth information from one image pair is a classical task in photogrammetry and also the minimal case of the well-known structure from motion problem. It refers to the concept that 3D structures can be recovered from the projected 2D motion field of a scene acquired with a moving sensor. Given the orientation between the images, depth can be reconstructed for each pixel using dense matching approaches like HIRSCHMÜLLER (2008), YAMAGUCHI et al. (2014) or CIGLA (2015).

Dense Matching has received much attention in recent years due to its wide range of deployment. The spectrum ranges from applications in the automotive industry to the creation of digital surface models and the digitization of cultural heritage. However, most of the approaches concentrate on matching images taken quasi-simultaneously (like in aerial photogrammetry) or in an environment with controlled illumination conditions. Consequently, there are no or only minor changes in illumination between such images. Only a few approaches address the problem that arises from the influence of different lighting conditions, despite the fact that research was already conducted on this topic back in the 1990s such as COX et al. (1995).

Early approaches tried to solve the problem by using more robust metrics for matching cost computation. There are a whole range of metrics such as Normalized Cross Correlation, Mutual Information, Rank and Census (ZABIH et al. 1994). In 2009, HIRSCHMÜLLER et al. investigated the influence of radiometric changes on the performance of these metrics. In their paper, they discuss the limitations and demonstrate that none of the examined approaches works particularly well, if there are bigger differences in illumination between the images.

To improve the robustness of dense matching against such cases, HEO at al. (2011) proposed a metric called Adaptive Normalized Cross-Correlation. Based on the observation that color consistency does not hold between stereo images in real scenes, they explicitly used the color formation model to adjust the matching cost function. Furthermore, the images were transformed using Log-Chromaticity Normalization beforehand. In contrast, ZHOU at al. (2012) and JUNG et al. (2013) rely on gradient based approaches. They assume that gradients are less sensitive to radiometric variations than the original pixel values and hence, they utilize gradient strength, orientation and distribution.

Another category of methods first transforms the images to an alternative representation before performing dense matching. KIM at al. (2014) proposes a transformation to the Mahalanobis distance transform space. It is defined as a collection of the Mahalanobis distances between the color of a pixel and the average color in this pixel's neighborhood. This is an attempt to separate the color values of the individual pixels from the influence of lighting. It is based on the assumption that the relation between the color and the average color is preserved under affine transformation. The work of MADDERN et al. (2014) proposes a transformation to an illumination invariant color space. They assume the sun as only light source and model it as a Planckian source. This approach works well for outdoor scenes illuminated by day-light. MOUATS et al.

(2015) proposes local frequency analysis which is based on Phase Congruency (KOVESI 1999). It combines information of edges and texture which is extracted from the image's frequency domain examined in a local context.

One issue that all of these approaches share is that they are based on assumptions that significantly restrict the scope of the method. Some can only deal with global, but not local lighting differences (ZABIH et al. 1994), others make specific assumptions about the type of light source (MADDERN et al. 2014) or the type of lighting change (KIM at al. 2014). As a result, these approaches are not suitable for building a reliable map, as it is based on images taken under different and quickly varying lighting conditions. Consequently, our approach is not based on any assumptions about the illumination, in order to be able to take all these images into consideration when creating the map.

3 Illumination Invariant Dense Image Matching

In order to be able to reconstruct depth information from a pair of images with highly varying illumination, we propose an approach based on the combination of multiple cost functions for dense matching. For this purpose, we assume that the orientation is known and the images have already been rectified.

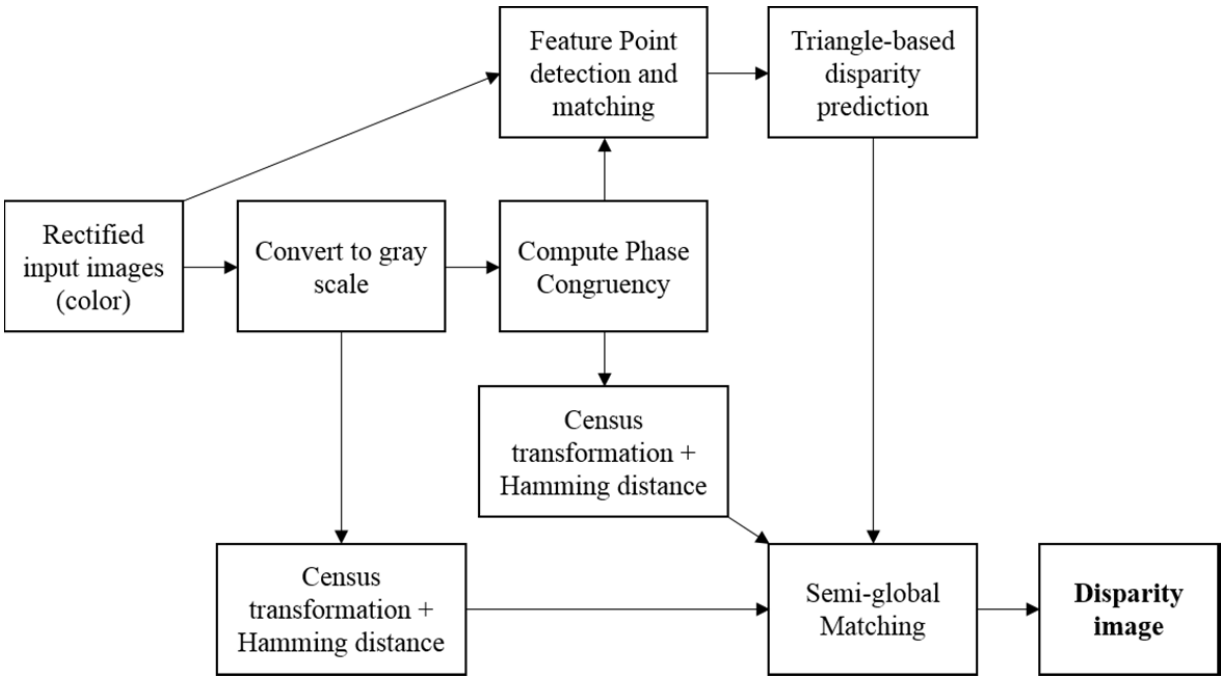

Fig. 1: Framework of the proposed dense image matching approach

As a basis we calculate the census transformation on the gray scale images as in ZABIH et al. (1994). The second element uses the Phase Congruency transformation of the images. Lastly, we use a triangle-based depth prediction, gained from a set of sparse feature matches, to support dense matching.

$$C(x,d) = \lambda_{Census} * C_{Census}(x,d) + \lambda_{PC} * C_{PC}(x,d) + \lambda_T * C_T(x,d) \quad (1)$$

The resulting cost function $C(x,d)$ is the weighted sum of the previously explained elements, with weights λ_{Census}, λ_{PC} and λ_T. It is computed for each pixel x and disparity d.

3.1 Phase Congruency

The Phase Congruency transformation is a concept developed by KOVESI (1999). It is based on the local energy model and shows the behavior of an image in frequency domain. The transformation can be used to detect and localize edges and corners and is invariant to changes in illumination and contrast. Consequently, the Phase Congruency remains constant for a scene even if the lighting conditions change.

Because of the Time-Frequency Uncertainty Principle it is not possible to exactly determine spatial position and frequency at the same time. Therefore, it is necessary to use an approximation to analyze the combined frequency-spatial space (SKARBNIK et al. 2010). Following RAGB et al. (2016) we use a bank of Log-Gabor filters for this purpose. Hence, the transformation can be computed over various scales n and orientations o with:

$$PC(x) = \frac{\sum_o \sqrt{(\sum_n (I(x) * M_{no}^e))^2 + (\sum_n (I(x) * M_{no}^o))^2}}{\sum_o \sum_n (A_{no}(x)) + \varepsilon} \quad (2)$$

where $I(x)$ is the input signal for which we use the gray scale image. M_{no}^e and M_{no}^o are the even and odd symmetric components of the Log-Gabor filters and ε is a small offset to prevent division by zero. The amplitude of the response A_{no} is defined by:

$$A_{no}(x) = \sqrt{(I(x) * M_{no}^e)^2 + (I(x) * M_{no}^o)^2} \quad (3)$$

We use 6 different orientations and 4 different scales, so we have a filter bank containing 24 different filters. We understand this as a compromise between accuracy and calculation effort. The impact on the cost function C_{PC} by increasing the number of filters was insignificantly small.

To compute the cost function C_{PC} both images are transformed with Phase Congruency, so an ordinary metric can be used to measure the similarity between image blocks:

$$C_{PC}(x,d) = F(I_{PC}^L(x), I_{PC}^R(x-d)) \quad (4)$$

where I_{PC}^L and I_{PC}^R are the left and right Phase Congruency images. As for the gray scale images, we use the Census transformation combined with Hamming distance for metric F.

3.2 Triangle-based disparity prediction

The third element of the matching cost function is based on triangle-based cost aggregation proposed by BULATOV et al. (2011). The basic idea is to build a triangle mesh based on a set of points for which the position is known in both images. The disparity for these points can then be directly computed from the difference of their positions in the two images. For all pixels within a triangle, the disparity is estimated by interpolating the disparity values of the triangle's vertices. This approach is comparable to methods based on image segmentation. However, it has the advantage that it is less prone to discretization errors and the number of disparity levels does not have to be set beforehand.

3.2.1 Sparse Feature Matches

In a first step, a set of points has to be specified which can be reliably found in both images. For this task, a feature detector and a descriptor are needed, which are robust against illumination changes. Research on feature detectors and descriptors has received much attention in recent years, but it was mainly focused on improving the reliability under viewpoint changes.

Nevertheless, there are some promising approaches like the TILDE detector proposed by VERDIE et al. (2015). Their approach is based on a piecewise linear regressor, which is learned beforehand. For training a series of images of the same scene taken under varying illumination conditions is used. On these images feature points are detected using SIFT (LOWE 2004). These points are introduced as positive samples for training. Negative samples are created by extracting patches at locations that are far away from the feature points.

In our work, we use the TILDE detector with a pre-trained linear regressor to detect feature points. In contrast to the original approach, we use a threshold on the regressor's value to filter for reliable points instead of keeping a constant number. As proposed by VERDIE et al. (2015) SIFT is used as descriptor. For the purpose of matching, a simple brute-force approach is utilized, which is improved by limiting the search space in vertical direction by a threshold ε. The set of potential matches for a feature point (x_L, y_L) detected in the first image is then defined by:

$$M_p(x_L, y_L) = \{(x_R, y_R) | y_L - \varepsilon \leq y_R \leq y_L + \varepsilon\} \tag{5}$$

where x_R, y_R are the coordinates of a potentially matching feature point in the second image. This approach is based on the assumption, that the input images are rectified and therefore, share a common image plane. Deviations in vertical directions are then caused by inaccuracies in the rectification process or movements in the scene only. Next, a distance is computed for every potential match in M_p:

$$M_d(x_L, y_L) = \{(x_R, y_R, \rho) |\ (x_R, y_R) \in M_p \cap \rho = F(x_L, y_L, x_R, y_R)\} \tag{6}$$

where F is a function that delivers the difference between two feature points based on their descriptors. From the set of potential matches a subset containing actual matches is extracted. For this purpose, we apply the ratio test (LOWE 2004) on the match with the minimal distance ρ for every feature point (x_L, y_L):

$$M(x_L, y_L) = \{(x_R, y_R, \rho_m) \in M_d(x_L, y_L) | \forall (x_R, y_R, \rho_n) \in M_d(x_L, y_L) \backslash (x_R, y_R, \rho_m) : \rho_m \leq \alpha * \rho_n\} \tag{7}$$

where α is a positive constant which regulates the strictness of the ratio test. For every feature point in the first image one match remains at most $|M(x_L, y_L)| \leq 1$. The final set of matches M is then defined as the union of all remaining feature point matches:

$$M = \bigcup_{x,y} M(x, y) \tag{8}$$

Furthermore, we detect the feature points not only on the color images, but also on the Phase Congruency images. The sets of detected feature points are quite different for both types of images. Hence, merging the sets of matched features leads to better coverage of the images and allows to create finer triangle meshes later on. Although the TILDE detector has been trained on

color images only, it works very well on the Phase Congruency images also. As descriptor SIFT is used as well.

Finally, the set of matched feature points is used to create a triangle mesh. For this purpose, the Delauney triangulation is used.

3.2.2 Triangular Interpolation

The approach of BULATOV et al. (2011) is based on the assumption that the non-occluded parts of a scene can be piecewise approximated by triangles. With a correct evaluation which triangles are nearly consistent with their covered surface, it is possible to approximate disparity values for all pixels within these triangles with high accuracy. However, this results in the difficulty to define constraints to filter triangles which include areas that cannot reasonably be approximated using a flat surface. For this purpose, we filter the set of triangles based on three policies: Size, variance of disparity values and variance of the color values of the included pixels. If the threshold of one or more policies is exceeded the triangle is excluded from the following processing steps.

In order to obtain disparity values for all pixels within the remaining triangles, the cost function introduced by BULATOV et al. (2011) is used as C_T:

$$C_T(x,d) = A(x,T) * D(d, d_{T,x}) \tag{9}$$

where $d_{T,x}$ is the interpolated disparity at pixel x within triangle T. The weight function A represents the reliability of the disparity in pixel x. It is maximal in the triangle's vertices and becomes smaller with increasing distance. Outside of the triangles boundaries A is zero.

$$A(x,T) = \hat{y} * exp\left(-\frac{g(x,T)}{\sigma}\right) \tag{10}$$

where the amplitude $\hat{y}$ and the descent parameter σ have to be greater than zero. g is the minimum distance from x to the vertices of triangle T. The function D evaluates the difference between the current disparity level d and the triangular prediction d_T in pixel x:

$$D(d, d_{T,x}) = \min\left(\frac{|d - d_{T,x}|}{d_0}, 1\right) - 1 \tag{11}$$

where $d_0 > 0$. Outside of the triangle's boundaries D is zero and accordingly, C_T is zero as well. Therefore, C_T only influences the value of the combined cost function C in areas which are covered by the triangle mesh. Based on the initial assumption this is reasonable, since it is not possible to make a reliable statement about the disparity in other regions.

3.3 Non-local optimization

In order to find the optimum disparity value from the combined cost function C for each pixel, Semi-Global Matching is used. According to HIRSCHMÜLLER (2008), the energy term E depends on the disparity image D and considers the combined matching costs C of (1) and a smoothness term:

$$E(D) = \sum_x (C(x, D_x) + \sum_{x' \in N_x} \lambda_1 T[|D_x - D_{x'}| = 1] + \lambda_2 T[|D_x - D_{x'}| > 1]) \tag{12}$$

The smoothness term adds a constant penalty λ_1 for all pixels x' in the neighborhood N_x of x, for which the difference in disparity is equal to one. For neighboring pixels with a difference in disparity bigger than 1, a constant penalty λ_2 with $\lambda_1 \leq \lambda_2$ is added. The introduction of a smoothness constraint with two different penalties is reasonable: Small changes in disparity typically indicate slanted surfaces and should be penalized less than sudden bigger changes. Also, bigger changes should be penalized uniformly in order to avoid oversmoothing.

The problem of finding the optimal disparity value for each pixel can now be solved by finding the disparity image D, which minimizes the energy $E(D)$. As mentioned in HIRSCHMÜLLER (2008) this is a NP-hard problem in 2D. Therefore, the smoothness constraint is approximated by evaluating only a specified set of paths instead of a global accumulation of all possible paths. Using (12) the path cost at pixel x and disparity d into direction r can be calculated by:

$$\begin{aligned} L_r(x,d) = \; & C(x,d) \\ & + \min\left[L_r(x-r,d), L_r(x-r,d\pm 1) + \lambda_1, \min_{d'}(L_r(x-r,d') + \lambda_2)\right] \\ & - \min_{d'}[L_r(x-r,d')] \end{aligned} \tag{13}$$

where $x - r$ is the previous pixel on path r. Since the minimum path cost of the previous pixel $\min_{d'}[L_r(x-r,d')]$ is independent from d, it can be used to bound the value of $L_r(x,d)$ by subtracting it from the whole term. Without this operation, the value of $L_r(x,d)$ would increase permanently, which could lead to very large values.

Finally, the costs for one disparity at pixel x are computed. For this purpose, the costs L_r are summed over all paths r. According to HIRSCHMÜLLER (2008), the number of paths must be at least 8 and should be 16 to ensure a good coverage of the image.

$$S(x,d) = \sum_r L_r(x,d) \tag{14}$$

The optimal disparity value at pixel x can then be determined by choosing the one with the lowest overall cost $\min_d S(x,d)$.

4 Results

In this section we first describe our experimental setup before we present both qualitative and quantitative results. For this purpose, we evaluate our approach on the KITTI stereo 2015 dataset (MENZE & GEIGER 2015) to demonstrate the performance on image pairs taken under similar lighting conditions. Furthermore, we examine the robustness of our dense matcher against different simulated changes.

4.1 Experimental Setup

For the purpose of evaluation, we use the quality measurement proposed in the KITTI stereo 2015 benchmark: The disparity of a pixel is classified as erroneous if the difference between the estimated and the ground truth value is bigger than 3px and bigger than 5% of the ground truth value. As first quality measure we use the percentage of stereo disparity outliers in the first frame, whereby only pixels for which an estimate is available in the calculated disparity image are taken into account. In order to be able to evaluate the density of the disparity images, the

coverage of the ground truth pixels for which a value is available in the estimation is also determined.

Since there is currently no benchmark for image pairs with different illumination conditions, we can only compare our approach with methods whose source code is freely available. Therefore, we compare to the following approaches: GEIGER et al. (2010), ZHANG et al. (2015), CIGLA (2015) and YAMAGUCHI et al. (2014). All of them are listed on the KITTI stereo 2015 leader board.

For all experiments we use the same set of parameter values. This set was obtained by random sampling of the parameter space with 1000 iterations and evaluation of the current configuration's performance on a subset of the data. For the weights of (1) we use $\lambda_{Census} = 0.4118$, $\lambda_{PC} = 0.3564$ and $\lambda_T = 0.2317$. The parameters in (10) and (11) are selected as $\hat{y} = 10$, $\sigma = 351$, $d_0 = 82$. And the penalties of semi-global matching in (12) are set to $\lambda_1 = 22$ and $\lambda_2 = 87$. The sizes of the Census transform masks are set to 9x11 on the gray-scale image and 5×5 on the phase congruency image.

4.2 Evaluation on image pairs with similar illumination

In order to be able to make a basic statement about the performance of our approach, we first test it on pairs of images taken under similar illumination conditions. For this purpose, we decided to use the training part of the KITTI Stereo 2015 dataset proposed by MENZE & GEIGER (2015). It contains 200 stereo images taken by a mobile mapping platform and, in contrast to the test part, the ground truth disparity images are available. The ground truth is semi-dense covering about 30% of the pixels.

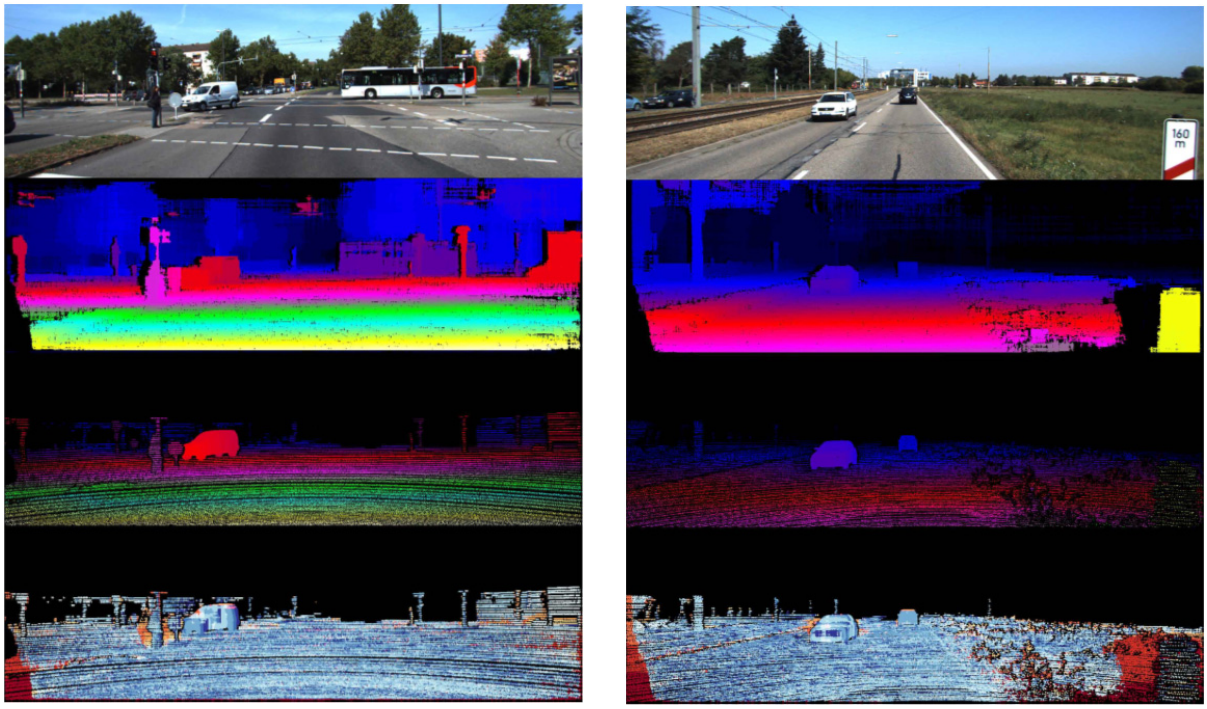

Figure 2: Qualitative results for our approach. In descending order: Left image, estimated disparity, ground truth disparity (from large disparities in yellow to small ones in blue), deviation between estimation and ground truth (from small deviations in white/light blue to large ones in orange/red).

To demonstrate the quality of the disparity images produced by our approach, Figure 2 shows two exemplary results. As one can see in the estimated disparity images (second from above) the contours are quiet sharp and only few artefacts appear. To view the quality of the results in relation to the state-of-the-art, Tab. 1 shows the averaged results for all examined approaches.

In contrast to most of the approaches on the benchmark's leader board, our approach is only semi-dense. Based on our goal to build a reliable map, it is important for us that the error contained in the estimated disparity image is as small as possible. Hence, we decided to not interpolate the disparity images for the moment to avoid the introduction of additional error. This decision results in a lower density and causes most of the differences between estimation and ground truth, which can be seen in the last images of Figure 2.

Tab. 1: Comparison with the examined approaches on the training part of the KITTI stereo 2015 dataset

Method	Error	Coverage
Elas Matcher (GEIGER et al. 2010)	5.71%	83.66%
Mesh Stereo (ZHANG et al. 2015)	7.80%	99.87%
REAF (CIGLA 2015)	13.13%	99.42%
SPS Stereo (YAMAGUCHI et al. 2014)	4.81%	99.97%
Ours	4.27%	89.00%

4.3 Evaluation on image pairs with simulated changes

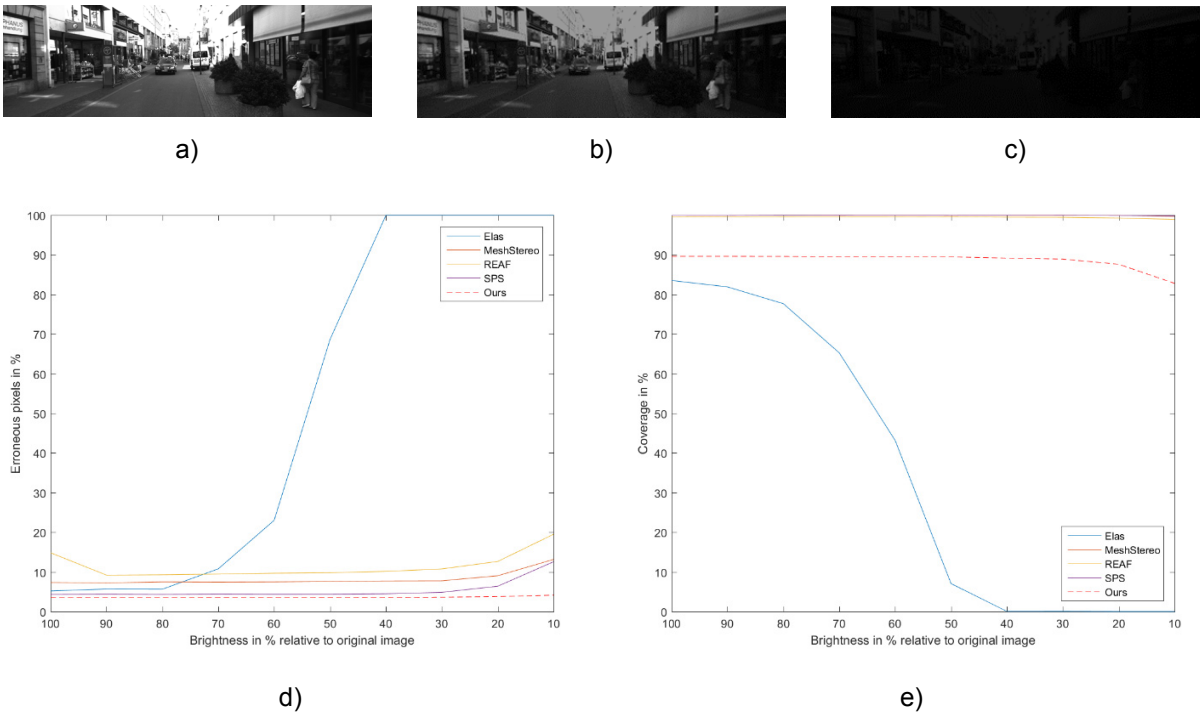

Fig. 3: Relative brightness decrease. (The brightness of the right image is reduced to the respective percentage.) a) - c) show the influence of the brightness reduction on the image at 100%, 50% and 10% of the original values. d) and e) show the effects of this change on the percentage of erroneous pixels and coverage.

In a second step, we evaluate the performance of our approach under varying conditions. For this purpose, we change the right image of a pair, while the left one is left untouched. In the following, the effects of relative brightness decrease, absolute brightness decrease and Gaussian noise are evaluated. For every type of change, the effect on the image, on the accuracy and also on the coverage is shown. The evaluation was performed on 10 different image pairs, with 10 different levels for relative brightness change, 15 different levels for absolute brightness change and 10 different levels for Gaussian noise.

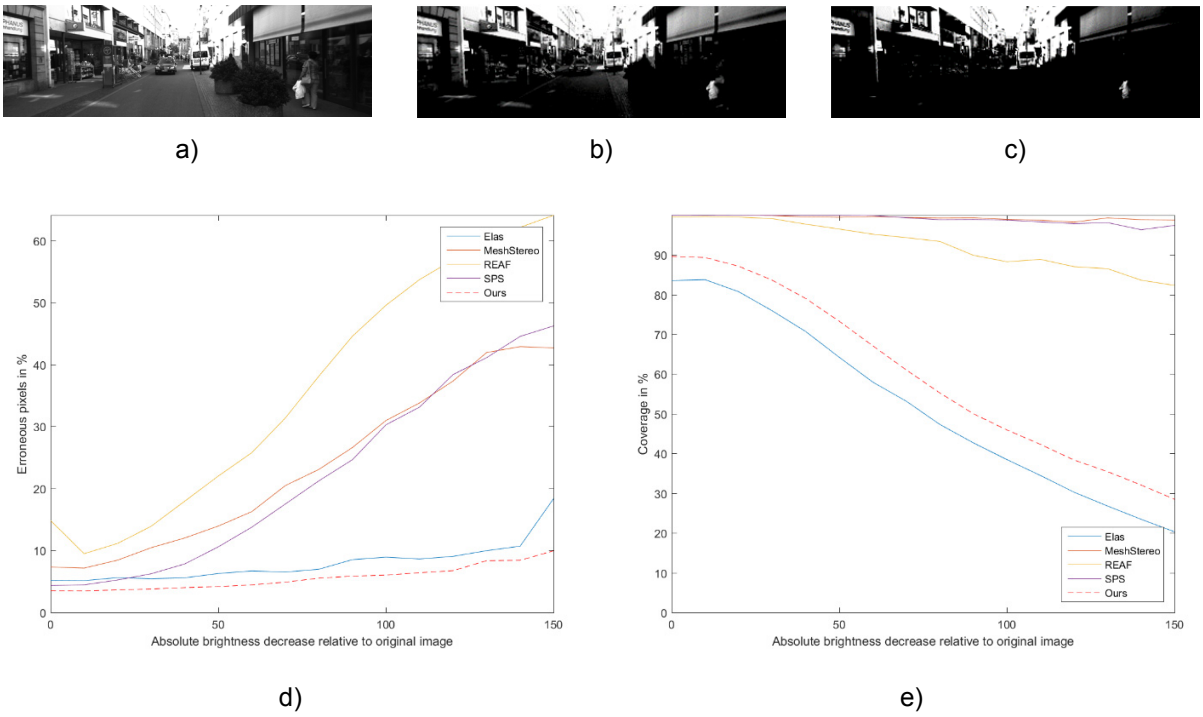

Fig. 4: Absolute brightness decrease. (The brightness of the right image is reduced by the respective value.) a) - c) show the influence on the image when the original values are reduced by 0, 80 and 150 grey levels. d) and e) show the effects of this change on the percentage of erroneous pixels and coverage.

As shown in Fig. 3 and Fig. 4 the percentage of erroneous pixels in the estimated disparity image increases much slower for our approach than for the others. In contrast, the coverage for our approach decreases comparatively quickly. This demonstrates that our approach is able to identify potentially wrong pixels up to a certain point. In ambiguous cases, the pixel is set to 0 instead of just choosing the value with the lowest energy.

A weakness of our approach can be seen in Fig. 5: The introduction of noise to one image leads to a high impact on error and coverage. Among other things, we want to address this problem in our future work.

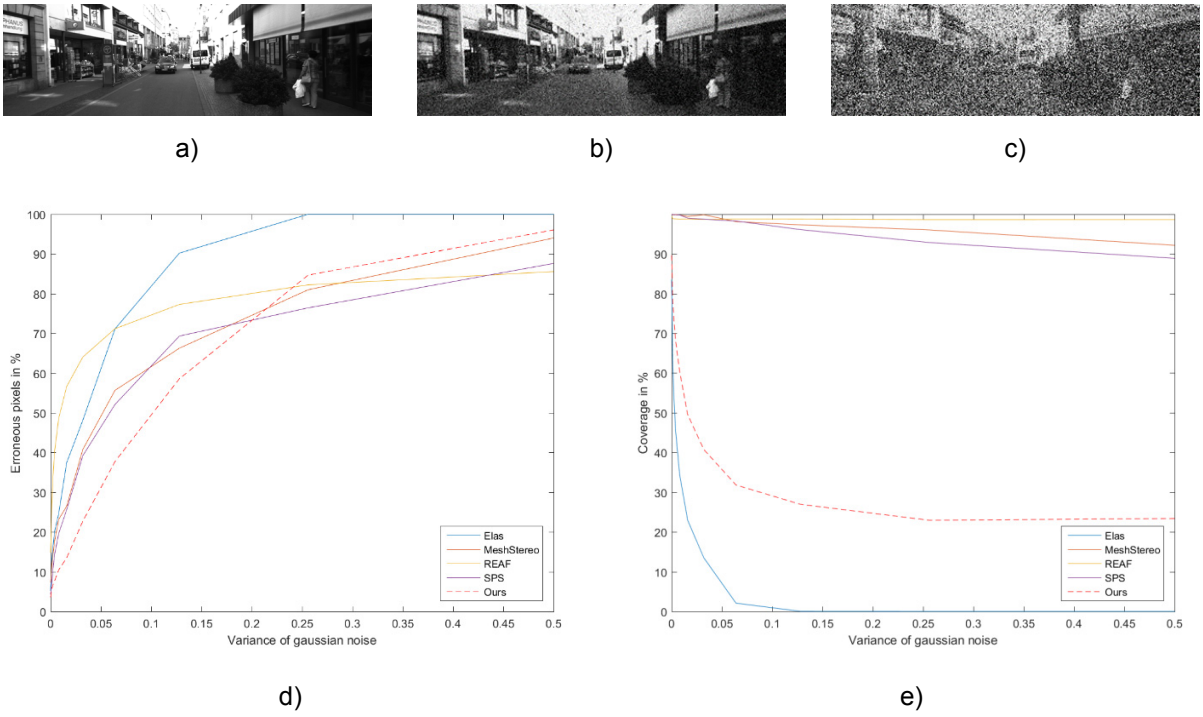

Fig. 5: Gaussian noise. (On the right picture Gaussian noise with respective variance is applied.) a) - c) show the influence on the image when applying Gaussian noise with a variance of 0, 0.25 and 0.5. d) and e) show the effects of this change on the percentage of erroneous pixels and coverage.

5 Conclusion

We have proposed an algorithm for dense image matching based on the combination of multiple cost functions to reliably reconstruct depth from an image pair. The performance of our approach was evaluated on the challenging KITTI stereo dataset, showing state-of-the-art results. Moreover, a set of images with simulated changes was used to demonstrate the robustness and the ability to detect potentially erroneous pixels. In future work, we want to investigate methods, which allow densification of the disparity image, without increasing the error. Furthermore, we believe that reliability measures for the single components within the cost function, could improve the performance under varying conditions.

6 References

BULATOV, D., WERNERUS, P. & HEIPKE, C., 2011: Multi-view dense matching supported by triangular meshes. ISPRS Journal of Photogrammetry and Remote Sensing, **66**(6), 907-918.

CIGLA, C., 2015: Recursive edge-aware filters for stereo matching. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 27-34.

COX, I.J., ROY, S. & HINGORANI, S.L., 1995: Dynamic histogram warping of image pairs for constant image brightness. Proceedings of the International Conference on Image Processing, **2**, 366-369.

GEIGER, A., ROSER, M. & URTASUN, R., 2010: Efficient large-scale stereo matching. Asian Conference on Computer Vision, 25-38.

HEO, Y.S., LEE, K.M. & LEE, S.U., 2011: Robust stereo matching using adaptive normalized cross-correlation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(4), 807-822.

HIRSCHMÜLLER, H., 2008: Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2), 328-341.

HIRSCHMÜLLER, H. & SCHARSTEIN, D., 2009: Evaluation of stereo matching costs on images with radiometric differences. IEEE Transactions on Pattern Analysis and Machine Intelligence, **31**(9), 1582-1599.

JUNG, I.L., SIM, J.Y., KIM, C.S. & LEE, S.U., 2013: Robust stereo matching under radiometric variations based on cumulative distributions of gradients, 20th IEEE International Conference on Image Processing (ICIP), 2082-2085.

KIM, S., HAM, B., KIM, B. & SOHN, K., 2014: Mahalanobis distance cross-correlation for illumination-invariant stereo matching. IEEE Transactions on Circuits and Systems for Video Technology, 24(11), 1844-1859.

KOVESI, P., 1999: Image features from phase congruency. Videre: Journal of computer vision research, **1**(3), 1-26.

LOWE, D.G., 2004: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, **60**(2), 91-110.

MADDERN, W., STEWART, A., MCMANUS, C., UPCROFT, B., CHURCHILL, W. & NEWMAN, P., 2014: Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles. Proceedings of the Visual Place Recognition in Changing Environments Workshop, IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, **2**, 3-10.

MENZE, M. & GEIGER, A., 2015: Object scene flow for autonomous vehicles. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3061-3070.

MENZE, M., HEIPKE, C. & GEIGER, A., 2015: Joint 3d estimation of vehicles and scene flow. ISPRS Workshop on Image Sequence Analysis (ISA), **8**.

MOUATS, T., AOUF, N. & RICHARDSON, M.A.,2015: A novel image representation via local frequency analysis for illumination invariant stereo matching. IEEE Transactions on Image Processing, 24(9), 2685-2700.

RAGB, H. & ASARI, V.K., 2016: Histogram of Oriented Phase (HOP): A New Descriptor Based on Phase Congruency. Electrical and Computer Engineering Faculty Publications. Paper 391.

SEIF, H. G. & HU, X., 2016: Autonomous Driving in the iCity-HD Maps as a Key Challenge of the Automotive Industry. Engineering, **2**(2), 159-162.

SKARBNIK, N., ZEEVI, Y.Y. & SAGIV, C., 2010: The Importance of phase in image processing. CCIT Report No. 773, 1-30.

VERDIE, Y., YI, K., FUA, P. & LEPETIT, V., 2015: TILDE: a temporally invariant learned detector. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5279-5288.

YAMAGUCHI, K., MCALLESTER, D. & URTASUN, R., 2014: Efficient joint segmentation, occlusion labeling, stereo and flow estimation. European Conference on Computer Vision, 756-771.

ZABIH, R. & WOODFILL, J., 1994: Non-parametric local transforms for computing visual correspondence. European conference on computer vision, Springer, Berlin, Heidelberg, 151-158.

ZHANG, C., LI, Z., CHENG, Y., CAI, R., CHAO, H. & RUI, Y., 2015: Meshstereo: A global stereo model with mesh alignment regularization for view interpolation. Proceedings of the IEEE International Conference on Computer Vision, 2057-2065.

ZHOU, X. & BOULANGER, P., 2012: Radiometric invariant stereo matching based on relative gradients. 19th IEEE International Conference on Image Processing (ICIP), 2989-2992.